Office for
National Statistics

Labour
Force
Survey

User Guide

VOLUME 1 – LFS BACKGROUND AND METHODOLOGY

BACKGROUND AND METHODOLOGY

Contents

.

SECTION 1 - THE HISTORY OF THE LFS IN THE UK

The Labour Force Survey (LFS) is a survey of households living at private addresses in the UK. Its purpose is to provide information on the UK labour market which can then be used to develop, manage, evaluate and report on labour market policies. The survey is managed by the Social Surveys division of the Office for National Statistics (ONS)[1] in Great Britain and by the Central Survey Unit of the Northern Ireland Statistics and Research Agency (NISRA) in Northern Ireland on behalf of the Economic Labour Market Statistics Branch (ELMSB) of the Department of Finance and Personnel.

For a more detailed description of the LFS and how it has developed, see:
- the August 2006 edition of Labour Market Trends *"Reflections on fifteen years of change in using the LFS: How the UK's labour market statistics were transformed by using the LFS", by Barry Werner (http://www.ons.gov.uk/ons/rel/lms/labour-market-trends--discontinued- /index.html).*
- the November 2013 release on "Forty years of change: UK's biggest survey marks its 40th birthday" *http://www.ons.gov.uk/ons/rel/mro/news- release/forty-years-of-change--uk-s-biggest-survey-marks-its-40th- birthday/uk-s-biggest-survey-marks-its-40th-birthday.html*

1.1 LFS 1973-1983

The first LFS in the UK was conducted in 1973, under a Regulation derived from the Treaty of Rome. The Statistical Office of the European Union (Eurostat) co-ordinates information from labour force surveys in the member states in order to assist the EC in matters such as the allocation of the European Social Fund. The ONS was responsible for delivering UK data to Eurostat until leaving the EU.

The survey was carried out every two years from 1973 to 1983 in the spring quarter (March-May) and was used increasingly by UK Government departments to obtain information which could assist in the framing and monitoring of social and economic policy. By 1983 it was being used by the Employment Department to obtain measures of unemployment on a different basis from the monthly claimant count and to obtain information which was not available from other sources or was only available for census years, for example, estimates of the number of people who were self-employed.

Published LFS estimates for 1973-1983 refer to the spring quarter and are available on a UK basis.

---

[1] Until 5 July 1995, the LFS was the responsibility of the Employment Department (ED). On that date ED was abolished and responsibility for the survey passed to the Central Statistical Office (CSO). On 1 April 1996, the CSO merged with the Office for Population Censuses and Survey (OPCS) to form the ONS which now has responsibility for the LFS.

1.2 ANNUAL LFS 1984-1991

Between 1984 and 1991 the survey was carried out annually and consisted of two elements:-

   (i)   A quarterly survey of approximately 15,000 private households, conducted in Great Britain throughout the year;

   (ii)   A "boost" survey in the spring quarter between March and May, of over 44,000 private households in Great Britain and 5,200 households in Northern Ireland.

Published estimates for 1984-1991 are available for the UK and are based on the combined data from the "boost" surveys and quarterly surveys in the spring quarters (Mar-May). The quarterly component of the 1984 to 1991 surveys were not published because the small sample sizes meant that the results were not robust. However, the quarterly survey proved to be invaluable in developmental terms, and in making early assessments of seasonality. A fuller description of the survey methodology used in this period is available in the annual results published by ONS (previously by OPCS) - see section 17 for details of these publications.

1.3 QUARTERLY LFS FROM SPRING 1992

In 1992 the sample in GB was increased to cover 60,000 households every quarter enabling quarterly publication of LFS estimates. Whilst it built on the annual survey, there were a number of differences which can be summarised as follows:

   (i)   panel design – from 1992 the GB survey was based on a panel design where a fifth of the sample each quarter is replaced and individuals stay in the sample for 5 consecutive waves or quarters. A shorter fieldwork period was also introduced which together with the panel nature of the survey led to slightly lower response rates.

   (ii)   sample design - the major difference was the introduction of an unclustered sample of addresses for the whole of Great Britain (the sample for Northern Ireland is similarly unclustered). This improved the precision of estimates particularly when making regional analyses. In the case of Scotland a very small bias arises from partial coverage of the population north of the Caledonian Canal. This area contains about five percent of the total population of Scotland.

   (iii)   additions to the sample - the inclusion of people resident in two categories of non-private accommodation, namely those in NHS accommodation and students in halls of residence. The students are included through the parental home.

In the winter of 1994/95 a quarterly Labour Force Survey was introduced to Northern Ireland. Each quarter's sample consists of approximately 3,000 household responses spread over five waves - 600 in each wave. A rotational pattern was also adopted, identical to that being operated in the GB LFS. Quarterly UK LFS estimates are available from winter 1994/95.

1.4 LFS QUARTERS

The quarterly LFS launched in 1992 in GB and in 1994 in NI operated on a seasonal quarter basis: March-May (Spring), June-August (Summer), September-November (Autumn) and December-February (Winter). The reasons for this were: -

(i)   Many activities associated with the labour market occur seasonally and follow the pattern of the school year. This was more the case when the LFS first started at which point more young people left school at Easter than in the summer;

(ii)   Easter can cause difficulty as it varies in timing between March and April – so ensuring that Easter is always covered by the same quarterly survey period avoids this problem.

The first results from the quarterly GB LFS, relating to spring 1992, were published in the LFS
Quarterly Bulletin (LFS QB) in September 1992 - that is, about 3½ months after the end of the survey period. From this date, the QB was the main source of LFS data. More timely results were presented in each quarter's ONS 'Labour Force Survey First Release' which provided key results about six weeks after the end of the survey period. Both the QB and the First Release presented GB estimates as Northern Ireland estimates were only available for the Spring quarters until Winter 94/95.

1.5 CALENDAR QUARTERS

In May 2006 the LFS moved to calendar quarters (CQ's). This means the micro data will no longer be available on a seasonal basis (spring – winter). The main reason ONS is moving to CQ's for the LFS is that it is an EU requirement under regulation[2]. Eurostat – the body responsible for the EU LFS – has a target structure for the survey with all Member States providing data on a CQ basis which will promote comparability across countries. In addition to conforming to the EU regulation, the switch from seasonal to calendar quarters will also enhance the comparability of the LFS with other quarterly surveys which are mostly conducted on a CQ basis. This is particularly relevant with respect to National Accounts.
The following table shows the resultant changes to the quarterly release of micro data.

Seasonal Quarters	Calendar Quarters (CQ's) – from May 2006
Winter (December to February)	Q1 = January to March (JM)
Spring (March to May)	Q2 = April to June (AJ)
Summer (June to August)	Q3 = July to September (JS)
Autumn (September to November)	Q4 = October to December (OD)

---

[2] Council Regulation (EC) No 577/98 and associated revisions.

This means the spring (March-May) questionnaire will move to the April-June questionnaire (Q2) and the June-August questionnaire will move to the July-September (Q3) and so on. Changes were also made to the interview weeks to align them to CQ's.

A note has been published in the June 2006 (*Labour Market Trends http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/lms/labour-market-trends--discontinued-/index.html*) which looks at the impact of the move to CQ's. There is also a CQ version of the Historical Quarterly Supplement (HQS) that was published on 17 May 2006 to coincide with the move. This will have historical data back to 1997 for certain quarters (mostly Q2 and Q4), so that users can look at trends based on CQ's. A partial series of micro data based on CQs has also been created covering the following periods: Q2 regional datasets 1997, 1999, 2001, and every quarter from then onwards.

A full back-series of micro data on a CQ basis has been produced.

1.6 EARNINGS FROM EMPLOYMENT QUESTIONS FROM WINTER 1992/93

Whilst questions in the LFS are continually being added, removed or modified, the major change to the early quarterly survey was the introduction of a section of earnings questions in GB from winter 1992/93 onwards. These questions were only asked of respondents receiving their fifth and final interviews, because of concerns that the questions might have an adverse impact on overall response rates. Results from these earnings questions were first published in the summer 1994 QB (in December 1994), and in the December 1994 *Employment Gazette*.

Earnings questions have been asked in the Northern Ireland LFS since the survey went quarterly in Winter 1994/5 but results were not weighted up until early 1998. LFS earnings data on a UK basis are available for each quarter from Winter 1994/5.

1.7 EARNINGS QUESTIONS FROM SPRING 1997

The LFS is an important source of earnings data, particularly for part-time workers. However, because earnings questions were initially only asked in wave 5 interviews, sample sizes were quite small and associated sampling errors tended to be relatively high. Work was done to test whether asking earnings questions in the first wave would lead to higher non-response in later waves, but no evidence was found to support this. So from Spring 1997 earnings questions were asked in both waves 1 and 5 in GB and NI, doubling the sample size and reducing sampling errors by about 30%. For more detail see 'Expanding the coverage of the earnings data in the LFS' in April 1998's *Labour Market Trends*.

1.8 MONTHLY PUBLICATION FROM WINTER 1997/8

A major public consultation on labour market statistics was conducted by ONS during 1997, resulting in a new integrated Labour Market Statistical Bulletin (LM SB), (previously called Labour Market Statistics First Release) first published in April 1998 (see February 1998 *Labour Market Trends* article 'Improved Labour Market Statistics'). The LM SB, which is published monthly, gives prominence to the ILO

measure of unemployment, as measured by the LFS over the administrative claimant count measure and draws together statistics from a range of sources to provide a more coherent picture of the labour market. The claimant count is not an alternative measure of unemployment.

LFS results in the LM SB are published on a UK basis, 6 weeks after the end of the survey period, and relate to the average of the latest three-month period. For the latest release see:
http://www.ons.gov.uk/ons/rel/lms/labour-market-statistics/index.html

Since April 1998, the Economic Labour Market Statistics Branch (ELMSB) of the Department of Finance and Personnel have published a Northern Ireland Labour Market Statistics Release to the same timetable as publication of the Labour Market Statistics First Release

1.9 Enhancements of the LFS in England, Wales and Scotland.

Since March 2000 extra respondents have been included as an annual enhancement to the sample size of the LFS. In March 2000 this was just for England, (and was known as the English Local LFS boost (ELLFS)) though was expanded to Wales (WLFS boost) in 2001/02 and Scotland (SLFS) in 2003/04.

These boost cases are interviewed annually for four years. More information on this can be found in the Volume 6 (APS) User Guide :
https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentand employeetypes/methodologies/labourforcesurveyuserguidance#labour-force-survey-lfs-user-guides.

The aim of the enhancements is to improve labour market information at a local level, as smaller sub-groups of the population can be looked at due to a larger sample.
When the results from the enhancements to the LFS in England, Wales and Scotland are combined, it is known as the Annual Local Area Labour Force Survey (ALAFS).

More information on the methodology behind the ELLFS can be found in the May 2000 and January 2002 issues of the Labour Market Trends

1.10 The Annual Population Survey

In 2004, a further improvement, the Annual Population Survey (APS), was introduced. The APS included all the data of the ALALFS, but also included a further sample boost in more urban areas of England – known as the APS(B) - aimed at achieving a minimum number of economically active respondents, in the sample, in each Local Authority District in England. The respondents included in the boost were not asked all the questions in the main LFS, see user guide Volume 6 for more information.

The first APS covered the calendar year 2004 rather than the ALALFS period of March to February. The ALALFS data was only published once a year, whereas the APS data is published quarterly, with each publication including a year's data,

In 2006, funding for the APS(B) was withdrawn, and so the structure of the Annual Population Survey reverted to the same as the ALALFS (that is, waves 1 and 5 of the quarterly LFS plus the ELLFS, WLFS, and SLFS). However, the name 'Annual Population Survey' has been retained, and the data continues to be published four times a year.

1.11 Changes introduced during and post COVID-19

9 March 2020: The public's heightened awareness of the coronavirus started to affect participation in the LFS.

17 March 2020: Wave 1 face-to-face data collection was suspended, in line with government guidelines, while systems were developed to allow interviewers to conduct telephone interviewing from their homes.

Telematching, which uses lookup information on telephone numbers associated with addresses (already used for respondents north of the Caledonian Canal) was extended to the rest of Great Britain to obtain additional telephone numbers for addresses in Wave 1.

LFS Waves 2 to 5 continued to be conducted via telephone where possible.

23 March 2020: Commencement of official UK lockdown measures.

Face-to-face interviewers restarted interviewing intermittently using telephone mode. Additional advance materials were prepared to allow respondents to contact interviewers.

30 March 2020: Telephone interviewing was rolled out fully to face-to-face interviewers.

20 April 2020: A new online portal was put in place to allow improved collection of respondent telephone details.

1 July 2020: Wave 1 sample size was doubled.

3 April 2021: Knock to Nudge commenced. Wave 1 sample reduced to 160% of pre pandemic level.

5 October 2021: Wave 1 sample reduced to 150% of pre pandemic level.

5 April 2022: Wave 1 sample increased to 155% of pre pandemic level.

July to September 2023: the issued sample returned to pre-pandemic level.

25 October 2023: reimplemented in-home interviewing for all Labour Force Survey (LFS) wave 1 and wave 2 cases.

October to December 2023: Work commenced to increase to sample size from 16,000 to 24,000. More information can be found in this [article](#) published in November 2023.

1.12 COVID-19 and Face to Face interviewing (and introduction of Knock to Nudge)

In March 2020, as a result of the coronavirus (COVID-19) pandemic, the LFS had to change the way it contacted people for initial interviews, from face-to-face interviewing to telephone-based. This change in method of initial contact has had an impact on both the level of response and the non-response bias of the survey, and consequently the survey estimates. Face to Face interviewing has been suspended since March 2020.

From April 2021, a field strategy referred to as 'Knock to Nudge' (KtN) was introduced on the LFS. This strategy involves interviewers visiting sampled addresses where no phone numbers could be obtained through either telematching or the online portal and encourage residents at the address to provide their phone number and arrange a telephone appointment. This field strategy proved to improve response rates as well as follow-up on those people that are otherwise 'harder to reach'.

1.13 COVID-19 and changes to the sample

Due to the drop in response rates at the onset of the pandemic the achieved sample for LFS main wave 1 for April to June 2020 only covered 69,000 individuals. To mitigate this problem, the issued sample size for the LFS main wave 1 was increased to maintain the achieved sample at the pre-pandemic level. This meant that the achieved LFS main wave 1 sample covered around 69,000 from July-September 2020 onwards. Due to the fact that only the main sample was topped up but not the boost sample, the achieved wave 1 sample for the overall number of economically active adults on the APS for June 2020 to July 2021 is 95,746, which is 36,904 (28%) below the target. This has continued since the with the onset of the pandemic and largely remained at that level since.

The issued sample size for the LFS wave 1 was reduced from April 2021 onwards due to the introduction of Knock to Nudge affecting the response rates positively and therefore still providing an achieved sample at pre-pandemic level.

1.14 COVID-19 and Weights

The mode change from largely face-to-face interviewing to exclusively telephone interviewing in wave 1 resulted in response bias as certain parts of the population that are naturally harder to reach where less represented in the achieved sample. Initial analysis observed a significant change to tenure, a household characteristic

that usually sees very little to now movement over time. An article [3]was published in October 2020 highlighting the observed changes in the estimates and the subsequent adjustment to the weighting methodology to account for this change. We introduced tenure weighting[4] which addressed some of the issues, but we acknowledged further work was needed and we committed to introducing further reweighting to improve LFS estimates.

Further analysis highlighted unexpected movement of other sub-groups of the population, such as the proportion of UK nationals increasing and non-UK nationals decreasing since the onset of the pandemic. Comparison with the HM Revenue and Customs (HMRC) payroll data with employees showed more subtle movements of these two groups. This led to the need for a further adjustment of the weighting methodology, outlined in an article published in March 2021, which used the HMRC Real Time Information (RTI) system as a data source to inform the required population growth adjustment.

Linked to the Migrant Worker Scan, the data helps infer the nationality of employees. This second adjustment was included in a reweighting exercise conducted between May to December 2021. which introduced an additional control in the weighting in relation to the structure of the population by country of birth.

The article[5] published in July 2021 explains the method being taken to reweight the LFS, including the introduction of the new controls. Revised weights were applied to LFS data from January to March 2020 onwards.

Following the release of new population estimates in 2023, an interim reweighting exercise was conducted and applied to the LFS person weights covering the periods from July-September 2022 onwards. Due to the way the reweighting was conducted, the new population totals were only applied to the LFS person level weight, not the income weight. From October-December 2023 onwards both the LFS person and income weight are based on the new population totals. An article was published in February 2024 providing further information on the method applied for this interim reweighting exercise. More information can be found in the following article **here**.

1.15 Multiple occupancy addresses since March 2020

When an interviewer identifies a concealed multi-household at a sampled address, a mechanism needs to be applied to ensure a random selection of one of the

---

[6] For the LFS: adult children living in halls of residence will be included at the parental address. For other ONS surveys a different definition exists The standard ONS instruction for defining a household states 'Adult children, that is, those aged 16 and over who live away from home should not be included at their parental address'.).

[6] For the LFS: adult children living in halls of residence will be included at the parental address. For other ONS surveys a different definition exists The standard ONS instruction for defining a household states 'Adult children, that is, those aged 16 and over who live away from home should not be included at their parental address'.).

[6] For the LFS: adult children living in halls of residence will be included at the parental address. For other ONS surveys a different definition exists The standard ONS instruction for defining a household states 'Adult children, that is, those aged 16 and over who live away from home should not be included at their parental address'.).

households at the sampled address. Due to the mode change at the onset of the pandemic, it was not possible for interviewers to identify such households anymore. For this reason interviewers have been instructed to interview the household they make contact with first. For consistency reasons, this approach has been continued despite the introduction of the KtN (see section 1.12) and later the face-to-face interviewing. As the occurrence of concealed multi-households is very small, the bias this selection process since the pandemic may have introduced is expected to be very small.

SECTION 2 - THE LFS IN NORTHERN IRELAND

The Northern Ireland Labour Force Survey is the responsibility of Economic Labour Market Statistics Branch (ELMSB) of the Northern Ireland Statistics and Research Agency (NISRA). Fieldwork is carried out by the Central Survey Unit, also of NISRA.

From 1973 - 1983, as in GB, the survey in Northern Ireland was conducted in alternate spring quarters. From 1984 - 1994 it was carried out annually. This annual survey consisted of 5,200 addresses drawn at random from the Rating and Valuation List - approximately 1% of private addresses in Northern Ireland. Over this period interviewing was conducted only in the spring, with no quarterly element. UK LFS estimates are available for Spring quarters from 1973-1994.

In the winter of 1994/95 a quarterly, Labour Force Survey was introduced to Northern Ireland. Each quarter's sample consists of approximately 3,000 household responses spread over five 'waves' - 600 in each wave. A rotational pattern was also adopted, identical to that being operated in the GB LFS. Respondents at 'wave' 1 are interviewed face-to-face with subsequent interviews at 'waves' 2-5 taking place, where possible, by telephone. Computer assisted interviewing has been used in the Northern Ireland Labour Force Survey since 1992. Quarterly UK LFS estimates are available from winter 1994/95.

Income questions have been asked in the Northern Ireland LFS since the survey went quarterly in Winter 1994/5 but results were not weighted up until early 1998. LFS income data on a UK basis is now available for each quarter from Winter 1994/5. From Spring 1997, the income questions in both the GB and NI LFS have been asked of respondents in waves 1 and 5, producing a larger sample size then when previously asked only of wave 1 respondents.

Since April 1998, the Economic Labour Market Statistics Branch (ELMSB) have published a Northern Ireland Labour Market Statistics Release to the same timetable as publication of the Labour Market Statistics Bulletin.

From January 2018 Northern Ireland have introduced a boost in the LFS sample. Within each quarter the number of households selected for participation in wave 1 was increased from around 780 to 1300. This increased sample will follow through each of the waves resulting in a boost in the total sample by quarter 1 2019.

SECTION 3 - SAMPLE DESIGN

The Labour Force Survey (LFS) is the largest regular social survey in the United Kingdom. The Office for National Statistics (ONS) conducts the survey in Great Britain, and its implementation in Great Britain is the responsibility of ONS' Social Survey Division, which works in close co-operation with ONS' Methodology Directorate. The Central Survey Unit of the Northern Ireland Statistics and Research Agency (NISRA) conducts the survey in Northern Ireland. The designs of both the Great Britain and Northern Ireland surveys are similar.

Though a quarterly survey, the design of the LFS and fieldwork procedures enable estimates of levels, such as the number of people in employment, to be produced for rolling three-monthly periods. Such estimates are published in the monthly Labour Market Statistics statistical bulletin.

This section of the User Guide examines the sampling procedures used in the LFS, the sample design has implications on the weighting used in the survey (see Section 10) and calculation of standard errors (Section 8). It also has close links with Fieldwork (Section 5), Non-response (Section 9) and Imputation (Section 12).The National Statistics Quality Review (NSQR) also contains some additional information: *http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-reviews/list-of-current-national-statistics-quality-reviews/nsqr-series--2--report-no--1/index.html*

3.1 TARGET POPULATION

3.1.1 Private Households

The target population of the LFS is based on the resident population in the United Kingdom. Specifically, the LFS aims to include all people resident in private households, resident in National Health Service accommodation, and young people living away from the parental home in a student hall of residence or similar institution during term time. (This latter group is included in the LFS sample specifically to improve the coverage of young people.)

Most recently in January to March 2024, the sample consists of around 23,000 responding (or inputed) households in Great Britain. Data from approximately 2,500 households (post boost) in Northern Ireland are added to this, representing about 0.3% of the NI population, allowing analysis of data relating to United Kingdom.

For most people, the meaning of residence at an address is unambiguous, and people with more than one address are counted as resident at the sampled address if they regard that as their main residence. The following are also counted as being resident at an address:

1. people who normally live there, but are on holiday, away on business, or in hospital, *unless* they have been living away from the address for six months or more;
2. children aged 16 and under, even if they are at boarding or other schools;
3. students aged 16 and over are counted as resident at their normal term-time address even if it is vacation time and they may be away from it.[6]

3.1.2 Communal Establishments and Non-Private Households

The LFS relates mainly to the population of the UK resident in private households, with the exception of NHS accommodation and student halls of residence. Therefore, this section of the User Guide has been included to assist users who wish to form a more complete picture of the UK population.

The 2001 and 2011 Population Census definitions state that communal establishments (CEs) provide managed residential accommodation[7]. Examples of CEs include residential care homes and university halls of residence. LFS outputs relate almost exclusively to the population living in private households, and, with a couple of notable exceptions, exclude most of the population living in CEs.

Of social surveys in the UK, the LFS is not alone in excluding CEs from its sampling frame; the Living Costs and Food Survey (LCF) and the Family Resources Survey do not sample from CEs either. Some departments (for example the Department of Health) do, however, occasionally conduct samples of sub-sets of the CE population.

At present, the decennial Population Census is the most reliable source of CE population data. In the past ONS has investigated options for surveying CEs on a more regular basis[8], but the main statistical obstacle remains the lack of a suitable, comprehensive and readily available sampling frame for all CEs.

Comparisons between LFS and Census estimates of the residents of communal establishments suggest that residents of CEs tend to differ from the rest of the population in terms of their demographic characteristics. The main differences are:
• there are proportionately more women in CEs
• the population is generally older in CEs, especially for women
• the economic activity rate is considerably lower amongst CE residents.

Table 3.1 provides estimates of the population resident in CEs in England and Wales from the 2021 Census.

---

[6] For the LFS: adult children living in halls of residence will be included at the parental address. For other ONS surveys a different definition exists The standard ONS instruction for defining a household states 'Adult children, that is, those aged 16 and over who live away from home should not be included at their parental address'.).

[7] See Population Definitions for 2001 Census (Census Advisory & Working Groups), Advisory Group Paper (99)04; and Final Population Definitions for the 2011 Census:
 http://www.ons.gov.uk/ons/rel/census/2011-census/population-and-household-estimates-for-the-united-kingdom/stb-2011-census--population-estimates-for-the-united-kingdom.html

[8] Communal Establishment Survey, Findings of the Pilot Stage: Summary Report, ONS (2009):

Table 3.1: Communal Establishments and their resident populations in England and Wales, as recorded by the 2021 Population estimates.

Communal establishment management and type	Number of residents
Medical and care establishment	368,118
Medical and care establishment:NHS	7,666
Medical and care establishment: NHS: General hospital	1,193
Medical and care establishment: NHS: Mental health hospital or unit (including secure units)	5,645
Medical and care establishment: NHS: Other hospital	828
Medical and care establishment:Local Authority	12,951
Medical and care establishment: Local Authority: Children's home (including secure units)	331
Medical and care establishment: Local Authority: Care home with nursing	2,911
Medical and care establishment: Local Authority: Care home without nursing	9,464
Medical and care establishment: Local Authority: Other home	245
Medical and care establishment:Registered Social Landlord or Housing Association	3,205
Medical and care establishment: Registered Social Landlord or Housing Association: Home or hostel	3,205
Medical and care establishment:Other	344,296
Medical and care establishment: Other: Care home with nursing	157,534
Medical and care establishment: Other: Care home without nursing	174,145
Medical and care establishment: Other: Children's home (including secure units)	3,287
Medical and care establishment: Other: Mental health hospital or unit (including secure units)	5,647
Medical and care establishment: Other: Other hospital	671
Medical and care establishment: Other: Other establishment	3,012
Other establishment	645,231
Other establishment: Defence	42,096
Other establishment: Prison service	65,327
Other establishment: Approved premises (probation or bail hostel)	1,527
Other establishment: Detention centres and other detention	348
Other establishment: Education	476,561
Other establishment: Hotel, guest house, B&B or youth hostel	15,048
Other establishment: Hostel or temporary shelter for the homeless	14,134
Other establishment: Holiday accommodation	1,427
Other establishment: Other travel or temporary accommodation	1,285
Other establishment: Religious	3,669
Other establishment: Staff or worker accommodation or Other	23,809
Establishment not stated	28,344

Total	1,041,693

Communal Establishments: The International Dimension

At an international level, European Union member states also exclude communal establishments, which is accepted by Eurostat due to the difficulties associated with including these in the sample. In the Labour Force Surveys of Australia, Canada and the USA, the sampling frames for the Labour Force Survey are designed to represent the civilian non-institutional population and therefore exclude:
- full-time members of armed forces,
- residents of institutions such as prisons and mental hospitals, and
- patients in hospitals or nursing homes who have been there at least 6 months.

In Australia some effort is made to include non-household residents using a list sample of non-private dwellings such as hotels and motels. The US equivalent of the LFS (the 'Current Population Survey') also attempts to include such people; the stratified sampling frame includes a 'group quarter' stratum containing those housing units where residents share common facilities or receive formal care.

3.2 SAMPLE DESIGN AND WAVE PATTERNS OF THE LFS

The LFS uses a rotational sampling design, whereby a household, once initially selected for interview, is retained in the sample for a total of five consecutive quarters. The interviews are scheduled to take place exactly 13 weeks apart, so that the fifth interview takes place one year on from the first.

We define Wave 1 to be the first quarter an address is selected, Wave 2 to be its second quarter in the selection, and so on. Therefore, Wave 5 is the last time that household will be interviewed for the main LFS. We stress here that it is the address that is selected for five quarters and not necessarily the particular people who live there. Therefore, it is possible to 'find' people new in the sample in Waves other than Wave 1, though the majority of people are first found in Wave 1. It is also possible for people to drop out of the sample before Wave 5 if they move to a different address.

The main reasons for use of a rotating sample design are:
- the precision of estimates of change over time is improved where there is overlap in the sample. Thus, better estimates of quarter-on-quarter and quarter on same-quarter-a-year-ago can be produced with this wave pattern;
- longitudinal data sets can be produced, which may be used for analysis of gross change (i.e. change in individuals' circumstances)

The same number of Wave 1 (new) addresses are selected each quarter. So, in any given quarter, about one-fifth of the addresses in the entire sample are in Wave 1, one-fifth in Wave 2, and so on. Thus, between any two consecutive quarters, about 80% of the selected addresses are in common. Figure 3.1 shows this pattern.

Figure 3.1: Wave patterns in the LFS.

	LFS Cohort 1	LFS Cohort 2	LFS Cohort 3	LFS Cohort 4	LFS Cohort 5	LFS Cohort 6	LFS Cohort 7	LFS Cohort 8	LFS Cohort 9	LFS Cohort 10	LFS Cohort 11	LFS Cohort 12
JM22	W5	W4	W3	W2	W1							
AJ22		W5	W4	W3	W2	W1						
JS22			W5	W4	W3	W2	W1					
OD22				W5	W4	W3	W2	W1				
JM23					W5	W4	W3	W2	W1			
AJ23						W5	W4	W3	W2	W1		
JS23							W5	W4	W3	W2	W1	
OD23								W5	W4	W3	W2	W1

The labelling of Cohorts in the diagram is arbitrary, and the same colour represent the same cohort of households.

Using JM22 as an example, we see that Cohort 5 (the dark green boxes), are having their Wave 1 interviews. In the same quarter, Cohort 4 will be having their Wave 2 interviews, Cohort 3 their Wave 3 interviews, Cohort 2 their Wave 4 interviews, and Cohort 1 their Wave 5 / final interviews.

Moving on one quarter to AJ22, and Cohort 5 are now having their Wave 2 interviews, Cohort 4 Wave 3 and so on. Cohort 1 is not interviewed in this quarter, and in its place, Cohort 6 has been selected for the first time and is on Wave 1 interviews.

Since each wave contains the same number of selected addresses, there is an 80% overlap between any two consecutive quarters. For example, between JM22 and AJ22, Cohort 2, 3, 4 and 5 are in common, Cohort 1 has been dropped and Cohort 6 is newly selected.

The LFS Waves in Great Britain were first created in the build-up period of the quarterly survey (autumn 1991 and winter 1991/92). Further details of this are reported in the 2009 (and earlier) editions of the LFS User Guide Volume 1.

The same pattern of waves is used in both Great Britain and Northern Ireland, but for the latter an additional sample, known as a booster, existed up to 2017. For the booster, 260 new Northern Ireland addresses (in addition to the usual new sample of 780 in 2017 and 650 prior to that) were added in Quarter 2 each year, and these were spread evenly amongst the five waves. Thus a booster address assigned to Wave 1 had four subsequent interviews, whereas one assigned to Wave 5 had no subsequent interviews. This booster has since been replaced with a permanent boost which began roll-out in 2018 with wave 1 and completed implementation with

quarter 1 2019. Including this boost a total on 1,300 new addresses are added to the sample each quarter.

3.3 SAMPLING FRAMES AND SAMPLE SELECTION

Four different sampling frames are used in the UK Labour Force Survey. Great Britain is split into two areas: south of the Caledonian Canal, comprising all of England, Wales and most of Scotland; and north of the Caledonian Canal in Scotland. Northern Ireland has its own sampling frame. A separate list of NHS accommodation in Great Britain was produced for the purpose of this survey.

The Wave 1 sample is selected by first ordering the sampling frames geographically, and then drawing the selection systematically (that is, with a fixed interval). The subsequent waves are not drawn from the frames; the Wave 1 selections are simply retained and become Wave 2 interviews in the next quarter, and so on.

For the most part, the LFS may be regarded as a single-stage sample of households each quarter, though changes made in 2010 (see Section 3.5) mean this is no longer strictly the case. The geographical ordering of the frame implicitly stratifies the sample, ensuring a geographic spread of addresses. Since all adults within a household are sampled, the person-level survey may be regarded (mainly) as a one-stage cluster sample of people, with the clusters (or primary sampling units) being the households.

We now look in more detail at each of the frames used, and how the selection of the Wave 1 sample is made. The information given refers to the number of addresses that are selected. Of course, not all of the addresses selected lead to a response, and we examine the number of responses in Section 3.7.

3.3.1 Sampling Households South of the Caledonian Canal in Great Britain

The sampling frame used for private households in Great Britain south of the Caledonian Canal is the Postcode Address File. The PAF is a computerised list, owned by Royal Mail, of all the addresses to which mail is delivered. The PAF is updated by ONS every six months.

The actual frame used for the LFS, and most other ONS social surveys, is the 'small users file', a sub-file of the complete PAF. 'Small users' are defined as delivery points which receive relatively few items of mail per day. This automatically excludes from the frame many businesses and other non-household institutions. However, the small users file still contains some non-private and non-residential (therefore ineligible) addresses, which cannot be identified prior to the interviewer making contact. Interviewers have instructions to exclude such institutions and classify them as ineligible.

The number of addresses selected from the PAF for Wave 1 each quarter is currently 25,800 since January-March 2024 – a number that has remained constant for many years up until the start of the COVID-19 pandemic but was subsequently adjusted several times to mitigate the drop in response due to the mode change (see section10). The selection process currently employed is as follows:

- The complete frame of delivery points is first ordered by Postcode, and within that by address.
- The sampling interval, k, (required for systematic sampling) is then calculated by dividing the total number of addresses (that is delivery points in England and Wales, and the multi-occupancy size marker in Scotland) by 25,800. This currently gives a 1-in-1586 Wave 1 quarterly sample size.
- A random start is chosen from {1, 2, …, k}, and that address and every k^{th} one after it are marked. This selection creates what is called the pre-sample.

To ensure no household is over-burdened, a Used Address File is maintained, such that an address used for sampling in any ONS social survey will not be sampled again for some two years or so after the final interview. To enable this, while also reducing any potential bias in small, local areas, the actual sample is then selected as follows:
- The number of marked addresses in each Postcode Sector (e.g. AB12 3..) is counted.
- A new systematic sample is then drawn separately for each Postcode sector from addresses not on the Used Address File. The sampling interval used in each Postcode sector is calculated so as to select the number of addresses required for that sector, as counted in the pre-sample.

All selected addresses (across all the five waves) are then allotted to pre-determined Interviewer Areas, and within those into weekly stints, 13 of which make up the quarter's interviews. More detail is given in Section 3.4.2.

3.3.2 Sampling Households North of the Caledonian Canal in Scotland

A different approach is taken for sampling north of the Caledonian Canal in Scotland. The canal runs from Corpach near Fort William on the west coast, through the lochs of Great Glen to Inverness on the east coast. The area to the north is sparsely populated, which means that interviewing a single-stage sample of addresses from the PAF face-to-face would be prohibitively expensive. An option of using a two-stage (clustered) sample design was considered, but the ultimate decision was taken to use a one-stage sample drawn from the telephone directory, along with telephone interviewing.

The sampling interval used on the main LFS sample south of the Caledonian Canal is used to determine the number of addresses to sample size north of the canal. Currently 80 addresses are selected for Wave 1 each quarter. Addresses are then selected systematically from the appropriate telephone directories, with the first one chosen with a random start, and following on in the directory from where the previous quarter's sample finished. Additional checks are made to ensure that the selected address is actually located north of the Caledonian Canal, and is not on the Used Address File.

The main disadvantage of sampling from telephone directories is the potential bias resulting from non-coverage of people not listed in the directory (e.g. those with no phone at all, a mobile phone only, ex-directory, or in a new-build property that is not yet listed). However, the alternative of a two-stage sample of addresses interviewed

face-to-face would still have led to large sampling errors and would also still incur high travel costs in the area.

3.2.3 Sampling Households in Northern Ireland

The sampling frame used in the Northern Ireland LFS is POINTER, which is the government's central register of domestic properties. It excludes commercial units. Land & Property Services (LPS) owns and maintains the register, and it is based on addresses held by the Ordnance Survey of Northern Ireland. It is updated two-to-three times a year, by LPS, the Northern Ireland District Councils, the Rates Collection Agency and other sources.

A similar selection procedure is used to that on the PAF, except the selection is made in one pass. Addresses that have been used recently for surveys are known as being 'flagged' in Northern Ireland, and these cannot be selected for the current period survey. As with the PAF, the frame is sorted geographically, ensuring a regional spread of sample addresses. The frame is sorted first by District Council, then by Ward and then by address.

The quarter's fieldwork is spread over three months, and a new sample is drawn every month. Prior to 2017, the total monthly sizes for the three months in a calendar quarter were 200, 250 and 200, giving a quarterly total of 650 Wave 1 (new) addresses. In 2017 the total quarterly sample was increase to 780 before a larger scale permanent boost was introduced in 2018. Again prior to 2018, 260 additional ('booster') new addresses were added to the sample in Quarter 2 of each year; these were spread equally across the five waves. This was replaced by a permanent boost in 2018 which, by full implementation in Q1 2019, had raised the monthly allocations to 396, 495, and 396 households. This currently represents a 1-in-1,900 Wave 1 sample of all domestic properties in Northern Ireland.

Allocation to interviewers is on a dynamic basis and takes into account total interviewing requirements and interviewer-availability.

3.3.4 Sampling NHS Accommodation

The sampling frame for NHS accommodation was specially developed for the Labour Force Survey. All district health authorities and NHS trusts were asked to supply a complete list of their accommodation (this accommodation mainly comprises what was once known as 'Nurses Homes', but the coverage is more extensive than that name implies)[9].

The proportion of addresses to sample is calculated by comparing the list with the PAF. Currently nine units of NHS accommodation in Great Britain are selected for Wave 1 interviews each quarter.

3.4 FURTHER NOTES ON SAMPLING

---

[9] Information was received from 417 out of the 455 authorities, trusts and teaching hospitals and the frame is not therefore complete. If the coverage of the frame is proportional to the coverage of authorities etc., then the frame contains 92 per cent of all NHS accommodation.

3.4.1 Multiple-occupancy addresses

Different sampling procedures exist at multiple-occupancy addresses, that is at those addresses at which more than one household resides or is likely to reside. Some of the more common examples include apartment blocks with just one front door, or a house which has been converted into flats. In Scotland, the Multiple-Occupancy marker on the PAF serves as a reliable guide to identifying the existence of multiple households behind the one front door. The marker is that used by the Post Office.

Within England, Wales, and sometimes still in Scotland and Northern Ireland, it is only when an interviewer first makes contact at the property that its multiple-occupancy structure becomes clear. In these cases, once the number of households present is established, prior to the pandemic just one of them was selected, at random, for interview. Since then the approach changed in that the first household the interviewer makes contact with is being interviewed. Section 3.5 gives more detail.

A slightly different scenario is that of the divided address. Again, typically, these are often one building that has been split into separate addresses. However, each address is listed separately on the PAF, but a marker is provided on all that belong to the one 'divided address'. In these cases, if it is the address with the highest address key (PAF unique identifier) within the building that is selected, the interviewer is asked to check there are no other addresses in existence in the building with the same postcode, that are not listed on the PAF. If there are, again it is just one that is selected for interview. This procedure attempts to ensure that all addresses in existence have a chance of selection.

3.4.2 Interviewer area allocations

We give some more details here of the way in which interviews are allocated to interviewers south of the Caledonian Canal. Further detail can be found in Section 5 of this volume of the LFS User Guide.
- The selected sample falls within 208 Interviewer Areas. These interviewer areas are split into "quotas", generally 2 in each interviewer area.
- For LFS fieldwork each quota is then divided into 13 stints, each stint containing roughly the same number of Delivery points (or MOs for Scotland). The Interviewer Areas are comprised of mainly two quotas, though there are some with one or three quotas (there are 318 quotas in England, 51 in Wales and 43 in Scotland).
- The 13 stints are randomly allocated to the 13 weeks of a quarter, and these are labelled 01 to 13. The Stint plus the week number form the quota number the same quota is covered by an LFS interviewer in the same week each quarter. Most interviewers cover two quotas . The design of the stinting is such that quotas are 'paired' so that an interviewer can be given 2 quotas of work they will be neighbouring. For example stint 901 and 902 are paired, so quotas 90101 and 90201 will be next to each other and any addresses which fall in those quotas will be interviewed in week one of the quarter.

- All postcodes are plotted on the boundary maps for the quotas, and the quota they fall in is held on the Sampling system. The systematic random sample of addresses selected for the quarter throughout the country is matched to it's quota on postcode to provide a list of addresses to be interviewed each week.
- A "Leap Week" is introduced periodically to re-align 'LFS quarters' (of 13 weeks) with calendar quarters, which gradually move out of alignment, as four quarters of 13 weeks give only 364 days, just short of a calendar year. The most recent LFS Leap Week was in October 2020 and was included to bring the LFS survey month into line with Eurostat regulations; the previous ones were in 2015, 2010 and in 2004. The Leap Week sees no LFS interviews take place (other than those left over from the previous week), and it is contained in neither the reference quarter before nor after.

3.4.3 Data collection modes

Most households are interviewed face-to-face at their first inclusion[10] in the survey and by telephone, if possible, at quarterly interviews thereafter. Respondents are encouraged to provide a telephone number and agree to interview in subsequent waves via the telephone.

Between January 2011 and December 2017, where a telephone number could be found and matched against an address selected in Wave 1, the household is first approach by telephone. This change was introduced from January 2011, and about 15% of addresses have their Wave 1 information collected by telephone. From January 2018 it was decided that these cases would be interviewed face to face in the first wave. Since 2014 a transformation programme is underway to develop a multi-mode online first Transformed LFS. ONS is planning to transition from the current to the new TLFS in 2024.

3.5 CHANGES TO THE LFS DESIGN IN 2010 AND ITS IMPLEMENTATION

In this section we note two changes made to the design in 2010 that mean the LFS samples in Great Britain and also in Northern Ireland are strictly no longer equal probability samples, although the effect of the changes is relatively small.

- The first change concerns multiple-occupancy addresses which are not separately identified as such on the frame. We first need to acknowledge that the PAF is a list of addresses, and that until an interviewer calls at that address, it is not known how many eligible households reside there. In most cases there is just the one household present at the listed address, but occasionally there will be more than one. Until the Q3 (July-September / JS) 2010 survey, all households at such an address were interviewed, and so all households had the same probability of being selected for the LFS (as do all adults within the household.)

---

[10] The small proportion of households sampled from North of the Caledonian Canal in Scotland are approached by telephone only.

From the Q3 2010 survey, only one household has been selected for interview where there was more than one present at the sampled address. The selection of that household used to be carried out randomly (i.e. by use of random numbers) up until the start of the COVID-19 pandemic. When the mode for wave 1 addresses had to change from face-to-face to telephone at the onset of the pandemic, the selection process for multi-occupancy addresses was adapted accordingly. Since then interviewers interview the household they make contact with first.. This change was initially introduced to help harmonise ONS social surveys. The effect of the change is that any such household now has a lower probability of selection, which is now reflected by it receiving a higher weight (see Section 10). The adaptation of the selection process with the start of the pandemic was considered to only have a minor impact on the quality of the achieved sample due to the small number of instances this process is applicable to.

This adjustment was first introduced in Q3 2010, and applied to Wave 1 households only. Thus all households in a multi-household address in Wave 1 in Q2 2010 or before continued to be follow-up for all five waves. Thus, the effect incrementally increases from Q3 2010 (Wave 1 only) to Q4 2010 (Waves 1 and 2) to Q3 2011 (all Waves).

- The second change in sample design was also introduced for the Q3 2010 interviews. If a household is found that has only adults aged 75+, then no further waves of interviews are conducted. This amendment had an immediate effect from its introduction, i.e. if a household of all 75+ occupants was found in any wave in Q2 2010, then no interview was conducted in Q3 2010 or any subsequent quarters.

The rationale behind this initiative, which makes considerable resource savings, is that such '75+' households tend to be stable in terms of their employment status. A corresponding change to the weighting of such households has been made (see Section 10), and thus a 75+ household found in Wave 1 (as is usually the case), now represents those in Waves 2 to 5 through a increased weight. The trade-off is, of course, that possible changes in employment status are missed, as would any change in the occupancy of the household over the next 12 months (for example if the 75+ households members moved out and another family moved in). Such 75+ households comprise about 8.6% of the Wave 1 sample.

3.6 SAMPLE DESIGN OF THE ANNUAL POPULATION SURVEY

3.6.1 Introduction to the APS and its Design

Volume 6 of the LFS User Guide details the Annual Population Survey (APS) and its data sets and data sources. However, as it is intrinsically linked to the LFS and its sample design, we also provide a summary here.

The design of the APS enables production of good-quality, annual estimates for relatively small areas of the United Kingdom on a rolling quarterly basis. Much of the data that comprise the APS data set come from the main LFS (Wave 1 and Wave 5

responses are pooled across four quarters); the remainder of the APS data set comes from boost / enhancement surveys in Great Britain. The APS data sets comprises data collected from the following three sources:

- Main LFS in the United Kingdom, Waves 1 and 5 only.
- The Local Labour Force Survey (LLFS) for England, Wales and Scotland. The LLFS is sometimes referred to as the LFS Boost, or occasionally, and somewhat erroneously, as the APS sample.
- The APS Boost, also known as the Annual Local Area Labour Force Survey (ALALFS) in England, which ran in years 2004 and 2005 only.

There is no boost sample in Northern Ireland, though we note that the sampling fraction in the main LFS in Northern Ireland is greater than that in Great Britain.

Within Great Britain, small areas for the boost samples are defined as:

- Local Authorities in London, of which there are 32.
- Local Education Authorities[11] elsewhere in England (at least up and including the design of the 2011 boost), of which there are 148.
- Local Authorities in Wales and Scotland, of which there are 22 and 32 respectively.

Each such area in Great Britain has a target number of interviews to achieve of Economically Active (EA) people (EA includes both employed and unemployed, according to the ILO definition). In some areas the target is achieved by the Main LFS itself, and no boost is required. (Recall here that the LFS sample is selected systematically from a geographically ordered list, thus the sample size in any given area is approximately proportional to its size.) In other areas, the Main LFS sample results in fewer achieved EA interviews than the target, and thus a boost is applied in that area. The targets were agreed some years ago by the bodies that fund the LLFS in England, Wales and Scotland, respectively the Department for Work and Pensions (DWP) and the Department for Business, Innovation and Skills (BIS), Welsh Government (WG) and Scottish Government (SG).

The APS, as its name implies, is an annual survey. Estimates are published each quarter, each being based on a rolling 4-quarter period. So as not to include data relating to the same household twice within any 4-quarter period, only Wave 1 and Wave 5 survey responses from the Main LFS are used in APS data sets. This is illustrated in Figure 3.2.

Figure 3.2: Main LFS wave patterns in the APS

---
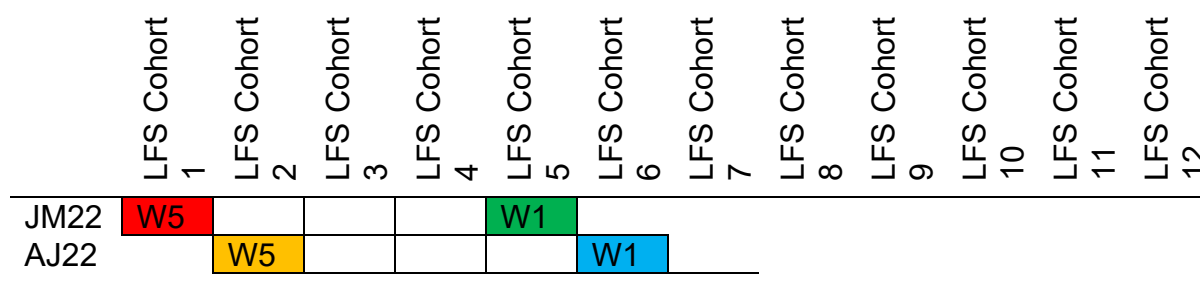
[11] The geographies used to define most areas in the England LFS Boost are currently under review; LEAs are no longer universally used, and a move to using UAs/LAs (or aggregates thereof) may result for future LFS Boosts.

JS22		W5				W1					
OD22			W5				W1				
JM23				W5				W1			
AJ23					W5				W1		
JS23						W5				W1	
OD23							W5				W1

- The Main LFS wave patterns and sample design are shown in Figure 3.1, but only data from Wave 1 and Wave 5 go on to form part of the APS data set.
- This pattern ensures no household (Cohort) will appear more than once in any rolling 4-quarter (i.e. rolling annual) data set. As an example, the top four rows (JM22 – OD22 inclusive) form the 2020 annual data set, and comprise data from Wave 5 interviews of Cohorts 1, 2, 3 and 4 and Wave 1 interviews from Cohorts 5, 6, 7 and 8.

The LLFS sample is designed with four annual waves (i.e. households sampled will be interviewed four times, each interview being a year apart), and the fieldwork is spread equally between the four quarters in the year. The wave design means that between any two consecutive years, 75 per cent of the LLFS sample is in common, and 25 per cent is replaced. This is shown in Figure 3.3

Figure 3.3: LLFS wave patterns in the APS

	LLFS Cohort 1	LLFS Cohort 2	LLFS Cohort 3	LLFS Cohort 4	LLFS Cohort 5	LLFS Cohort 6	LLFS Cohort 7
JM21	W4	W3	W2	W1			
AJ21	W4	W3	W2	W1			
JS21	W4	W3	W2	W1			
OD21	W4	W3	W2	W1			
JM21		W4	W3	W2	W1		
AJ21		W4	W3	W2	W1		
JS21		W4	W3	W2	W1		
OD21		W4	W3	W2	W1		
JM22			W4	W3	W2	W1	
AJ22			W4	W3	W2	W1	
JS22			W4	W3	W2	W1	
OD22			W4	W3	W2	W1	
JM23				W4	W3	W2	W1
AJ23				W4	W3	W2	W1
JS23				W4	W3	W2	W1
OD23				W4	W3	W2	W1

The APS data set is formed by data from the Main LFS and the LLFS, which has an annual, 4-wave pattern.

The first LLFS interview is by a face-to-face interviewer or on the phone where the respondent prefers it, and subsequent interviews are by telephone where the respondent agrees.

3.6.2 Design of the Local Labour Force Survey in England, Scotland and Wales

The LLFS is stratified by local area, with the areas defined in Section 3.6.1. The boost sample size has been selected as required to achieve the target number of EA interviews but, of course, this may not happen in reality due, for example, to changing response rates. The boost sample in each local area is reviewed each year. The process for determining any adjustments to the boost size in each area is summarised as follows:

1. The achieved number of EA interviews from Waves 1 and 5 of the main LFS sample size for the previous year is obtained. If this exceeds the target, no boost is required.
2. For other areas, the combined main LFS Wave 1 and 5 sample size, plus existing LLFS size is considered. Based on assumptions about response rates and wave-to-wave attrition, a projected number of achieved interviews is made for the forthcoming three years, by which time the previous year's sample size would apply to all the waves. If that projection is within a given tolerance of the target (currently set at 10%), no change to the boost sample is made for the coming year; if it is outside the tolerance, the boost sample size is adjusted (increased or decreased) for the forthcoming year is made, which brings the projection into line with the target in three years' time.

As the method for determining the boost size is based on actual, and recently achieved interview numbers, changes in response rates are implicitly taken into account, although there is a lag in them being reflected in the new sample sizes. The fall in response observed over recent years has resulted in the overall boost sample size increasing.

The LLFS's stratified design is reflected in the way the APS data are weighted; the local area of the household determines its design weight. APS weighting is described in Section 10.5.

3.7 SAMPLE SIZE INFORMATION: A SUMMARY

This section contains details of the sample sizes obtained in 2018 and 2019.

We give summary information about the number of delivery points selected, the number of eligible households, and response information by Wave (this information is only given for the Great Britain sample). Noting that the final data sets made available contain both actual responses, and imputations, information is given in this section about the number of imputations. For further information, we suggest the following sections of the User Guide:
- Response rates over time: Section 5
- Proxy responses (included within all responses in this section): Section 5
- Imputation: Section 12

3.7.1 Main Labour Force Survey Sample

Size of selected sample

As described in Section 3.3, the number of addresses selected for Wave 1 each quarter is as follows, and changes little over time. Except for the fluctuations in sample size and response rate due to the COVID pandemic:

- 25,800 household addresses from the PAF for Great Britain south of Caledonian Canal. Of these, 22,193 are in England, 1,299 in Wales and 2,308 in Scotland (These figures will vary very slightly each quarter), proportionally reflecting the number of delivery points (and multiple-occupancy markers in Scotland) in each country.
- 80 phone numbers matched for household addresses north of the Caledonian Canal.
- Around 1,300 household address in Northern Ireland.
- Nine units of NHS accommodation in Great Britain.

Thus, in any one quarter, a total of about 27,189 addresses are newly-selected in the UK for the main LFS (excluding the Northern Ireland boosters).

Since there are five waves in any given quarter, the total number of addresses selected in a given quarter is about 5 x 27,189 = 135,945. Of course, not all of these addresses selected first will be eligible, respond, or agree to take part in subsequent interviews / waves.

Size of responding sample

The sample is currently made up of approximately 23,000 responding UK households and 50,000 individuals per quarter (as of January to March 2024). Summary information on the number of households and people in the LFS in Great Britain and the UK is shown in Table 3.2.

Table 3.2: Household and person responses (including imputations and NHS accommodation)

		Numbers			Quarter-on-previous quarter change (%)			
Period	GBHH	GB People	UK HH	UK People	GBHH	GB People	UK HH	UK People
JM20	30,585	71,105	33,329	77,903	-7.3	-7.4	-7.3	-7.3
AJ20	27,798	64,040	30,093	69,733	-9.1	-9.9	-9.7	-10.5
JS20	28,992	64,765	30,923	69,425	4.3	1.1	2.8	-0.4
OD20	31,808	70,341	33,685	74,832	9.7	8.6	8.9	7.8
JM21	34,987	77,192	37,098	82,015	10.0	9.7	10.1	9.6
AJ21	37,529	82,715	39,799	87,904	7.3	7.2	7.3	7.2
JS21	38,208	84,223	40,940	90,453	1.8	1.8	2.9	2.9
OD21	35,335	28,084	38,136	30,735	-7.5	-66.7	-6.8	-66.0
JM22	33,407	28,235	36,236	30,929	-5.5	0.5	-5.0	0.6
AJ22	31,502	24,607	34,552	27,491	-5.7	-12.8	-4.6	-11.1
JS22	29,129	23,957	31,820	26,500	-7.5	-2.6	-7.9	-3.6

OD22	26,173	21,586	28,879	24,177	-10.1	-9.9	-9.2	-8.8
JM23	23,957	20,015	26,712	22,655	-8.5	-7.3	-7.5	-6.3
AJ23	21,855	17,470	24,572	20,060	-8.8	-12.7	-8.0	-11.5
JS23	17,824	14,180	20,453	16,645	-18.4	-18.8	-16.8	-17.0
OD23	17,822	15,182	20,324	17,544	0.0	7.1	-0.6	5.4
JM24	20,685	17,046	23,143	19,378	16.1	12.3	13.9	10.5

Source: Table 1 (Table 3 after JM20) of LFS Performance and Quality Monitoring Reports: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeet ypes/methodologies/labourforcesurveyperformanceandqualitymonitoringreports

The number of cases in Wave 1 is larger than the number in Wave 5 because of attrition in the sample and the sampling scheme now implemented for over-75 households. Wave information about the Great Britain LFS sample is shown in Table 3.3

Table 3.3: Wave-specific household responses and response rates for the LFS sample in Great Britain

		Wave 1		Wave 2		Wave 3		Wave 4		Wave 5		Total	
		#	%	#	%	#	%	#	%	#	%	#	%
JM22	Eligible, of which:	23,792	100	22,724	100	23,911	100	23,825	100	30,739	100	124,991	100
	responded	8,515	35.8	6,992	30.8	6,380	26.7	5,791	24.3	5,742	18.7	33,420	26.7
	imputed	-	-	1911	8.4	1287	5.4	1114	4.7	870	2.8	5,182	
AJ22	Eligible, of which:	24,475	100	22,451	100	22,604	100	23,891	100	23,801	100	117,222	100
	responded	7,813	31.9	7,666	34.1	5,416	24	5,453	22.8	5,170	21.7	31,518	26.9
	imputed	-	-	2514	11.2	1576	7	1549	6.5	1268	5.3	6,907	
JS22	Eligible, of which:	24,260	100	23,177	100	22,330	100	22,566	100	23,874	100	116,207	100
	responded	7,650	31.5	6,912	29.8	5,642	25.3	4,342	19.2	4,594	19.2	29,140	25.1
	imputed	-	-	2097	9	1370	6.1	941	4.2	773	3.2	5,181	
OD22	Eligible, of which:	24,264	100	23,541	100	23,503	100	22,639	100	22,764	100	116,711	100
	responded	24,264	100	23,541	100	23,503	100	22,639	100	22,764	100	116,711	100
	imputed	-	-	1915	8.1	1159	4.9	957	4.2	565	2.5	4,596	
JM23	Eligible, of which:	24,304	100	23,779	100	23,779	100	23,698	100	22,772	100	118,332	100
	responded	6,970	28.7	5,321	22.4	4,163	17.5	3,862	16.3	3,641	16	23,957	20.2
	imputed	-	-	1814	7.6	873	3.7	730	3.1	526	2.3	3,943	
AJ23	Eligible, of which:	24,317	100	23,806	100	23,873	100	23,825	100	23,732	100	119,553	100
	responded	6,794	27.9	5,016	21.1	3,506	14.7	3,330	14	3,209	13.5	21,855	18.3
	imputed	-	-	2035	8.5	1009	4.2	737	3.1	604	2.5	4,385	

Source													
JS23	Eligible, of which:	15,618	100	23,983	100	23,961	100	23,906	100	23,813	100	111,281	100
	responded	4,771	30.5	4,760	19.8	3,042	12.7	2,579	10.8	2,673	11.2	17,825	16
	imputed	-	-	1942	8.1	802	3.3	527	2.2	374	1.6	3,645	
OD23	Eligible, of which:	15,564	100	14,887	100	24,041	100	23,978	100	23,895	100	102,365	100
	responded	5,284	34	4,210	28.3	3,400	14.1	2,618	10.9	2,317	9.7	17,829	17.4
	imputed	-	-	265	8.5	687	2.9	440	1.8	250	1	2,642	
JM24	Eligible, of which:	24,073	100	14,927	100	14,888	100	24,030	100	23,970	100	101,888	100
	responded	7,880	32.7	4,506	30.2	3,080	20.7	2,881	12	2,344	9.8	20,691	20.3
	imputed	-	-	1660	11.1	858	5.8	694	2.9	427	1.8	3,639	

Source: Table 3 (Table 5 after JM20) of LFS Performance and Quality Monitoring Reports:

https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyperformanceandqualitymonitoringreports

Note that the eligible number of households may increase from one quarter to the next, for example if a household is found in Wave 2 in what was an unoccupied address in Wave 1.

Responses include full and partial response, but exclude imputed households. The sum of 'responded' and 'imputed' is consistent with the 'GB HH' column in Table 3.2 (noting minor discrepancies due to rounding: up to +/-5)

Eligible households, which didn't respond or were not imputed, may be regarded as other non-response.

3.7.2 Annual Population Survey

The sample size of the LLFS in England, Scotland and Wales does not remain constant from year-to-year, unlike that of the main LFS. Over recent years, the size of the LLFS has increased, reflecting decreasing response rates. The size of the 2023 (JD) APS data set of responses is given in Table 3.4.

The 2023 data set consists of **146,419** responding or imputed people, from 72,524 households. Of responding households in Great Britain in the data set, 52.0% came from LFS (the rest from LLFS), and of households in the UK, 55.0% (the higher proportion resulting from no boost in Northern Ireland). In terms of people, in Great Britain 54.0% are from the LFS data set, and of the UK, 53.0% are from the LFS.

Table 3.4: Households (responding or imputed) in the APS data set for JD23 (January 2023 to December 2023)

Source	Wave	England	Wales	Scotland	Northern Ireland	GB	UK

Source	Wave	England	Wales	Scotland	Northern Ireland	GB	UK
LFS	All waves of which:	**30,745**	**1,841**	**3,073**	**4,396**	**35,659**	**40,055**
	1	20,622	1,194	2,003	2,427	23,819	26,246
	5	10,123	647	1,070	1,969	11,840	13,809
LLFS	All waves of which:	20,544	6,342	5,583		32,469	32,469
	1	7,541	2,024	1,808		11,373	11,373
	2	4,896	1,484	1,459		7,839	7,839
	3	4,252	1,501	1,198		6,951	6,951
	4	3,855	1,333	1,118		6,306	6,306
Total		**51,289**	**8,183**	**8,656**	**4,396**	**68,128**	**72,524**

Source: direct analysis of the JD23 APS person- level data set.
The rows for the main LFS in GB are consistent with Table 3.3, with some minor discrepancies (up to +/- 2 HHs) due to rounding.

Table 3.5: Persons (responding or imputed) in the APS data set for JD23 (January 2023 to December 2023)

Source	Wave	England	Wales	Scotland	Northern Ireland	GB	UK
LFS	All waves of which:	**64,710**	**3,810**	**5,981**	**10,021**	**74,501**	**84,522**
	1	43,727	2,526	3,925	5,417	50,178	55,595
	5	20,983	1,284	2,056	4,604	24,323	28,927
LLFS	All waves of which:	**39,890**	**12,041**	**9,966**		**61,897**	**61,897**
	1	16,268	4,266	3,471		24,005	24,005
	2	9,059	2,765	2,499		14,323	14,323
	3	7,745	2,742	2,113		12,600	12,600
	4	6,818	2,268	1,883		10,969	10,969
Total		**104,600**	**15,851**	**15,947**	**10,021**	**136,398**	**146,419**

Source: direct analysis of the JD23 APS person-level data set.

Note on Tables 3.4 and 3.5:
The wave patterns used in the main LFS and the LLFS mean that:
- the LFS Wave 1 households here were first interviewed in 2023, whereas the Wave 5 households here were first interviewed in 2022.
- the LLFS Wave 1 households were first interviewed in the LLFS in 2023, Wave 2 here were first interviewed in 2020, Wave 3 in 2021 and Wave 4 in 2020.

SECTION 4 - THE QUESTIONNAIRE

4.1 MANAGEMENT OF THE LFS QUESTIONNAIRE

The questionnaire content is determined by ONS. ONS are responsible for identifying, in conjunction with other government departments, needs for new questions or changes to existing questions (e.g. changes in legislation or new government employment programmes) and for determining priorities, given the constraint of interview length.

A number of other Government Departments also sponsor LFS questions, including the Department of Transport (travel to work) and the Health and Safety Executive (accidents at work).

Discussions between ONS and other Government Departments on the questionnaire content for all the four quarters used to follow an annual cycle. Typically, the Labour Market Division in ONS and other Government Departments would submit in December an outline for requirements for the survey beginning 13 months from then to the Social Survey Division in ONS. Initial discussions were carried out at the start of the year and a package of questions were tested to see that they are acceptable and understood by respondents. A decision will be made to see if there is a need for cognitive interviewing (to pilot the questions) before the Dress Rehearsal (a further round of testing). The Dress Rehearsal, which usually took place around July (though this did vary), tests whether potential new questions fit in well with the overall questionnaire. However before any new questions could be added to the questionnaire, room needs to be found to avoid the questionnaire getting any longer. By September, the broad content for the following year would be agreed. Final agreement from the LFS Steering Group was normally required in October. The new questionnaires then went live a few months later, starting with the January to March quarter.
Do to the ongoing transformation work on the Transformed Labour Force Survey (TLFS) which is going to replace the current Labour Force Survey, a decision was made in 2022 to focus development resource on the TLFS. Therefore no major changes have been implemented into the questionnaire since then.

Throughout, the interests and priorities of other government departments are taken into account via the inter-departmental LFS Working Group, which brings together departments with particular interests in LFS data every two years.

4.2 QUESTIONNAIRE DESIGN AND STRUCTURE

The questionnaire comprises a "core" of questions which are included in every quarter of the survey, together with "non-core" questions which are not asked every quarter. These "non-core" questions provide information that is needed less frequently. Some "non-core questions are only asked in one or two quarters per year, for example, the majority of the questions on a respondents employment pattern are only asked in the second quarter. Other "non-core" questions do not appear every year, but are included in the survey every 2 or 3 years. For example, questions on regional mobility are asked every 3 years.

Some questions in the core are only asked at the first interview (wave 1) as they relate to characteristics that do not change over time (e.g. sex, ethnicity, country of birth and nationality). There have also been some more wave 1 questions and a wave 1 weight (EWEIGH**[12]) added to the Government cuts of the JD APS person datasets (see user guide volume 6 for more information).

Since spring 1997, a section on earnings from employment, has been asked in respondents first and fifth interviews (prior to that it was asked only in the fifth interview). The earnings data are processed along with the rest of the data each quarter but are weighted separately.

---

[12] Where ** denotes the year that the weight was published.

SECTION 5 - FIELDWORK

5.1 THE CONDUCT OF FIELDWORK

Face-to-face and telephone interviewing

LFS fieldwork is carried out by the Labour Force Survey interviewing community which is comprised of both face-to-face interviewers, who work from their homes, and by telephone interviewers, who work in a centralised Telephone Operations Unit in Titchfield, Hampshire, where close supervisory control over the conduct and quality of interviews can be maintained. Interviewer managers regularly accompany face-to-face interviewers to ensure that standard procedures are being implemented and the instructions issued to interviewers on the interpretation and coding of responses are being followed. Many of the interviewers work on a part-time basis and there is some spare capacity to allow for cover for sickness and other absences.

The majority of first interviews (wave 1) at an address used to be carried out face-to-face, except those North of the Caledonian Canal (see section 3) and those where the telephone number can be matched to the address (Approximately 160 addresses per quarter). If the respondent agrees to it, recall interviews are carried out by telephone. Since the COVID-19 pandemic, when at first a telephone only approach was followed, the take-up of face-to-face interviews in wave 1 is lower. Where previously the vast majority of wave 1 addresses were interviewed face-to-face, around half of addresses take up the offer of face-to-face interviewing. The remainder prefer to be interviewed over the telephone.

Number of interviewers

As mentioned above, the interviewing community for the LFS consists of both face-to-face and telephone interviewers. In March 2024 there were approximately 477 interviewers working in the field and 148 in the telephone operations.

Timing of interviews

The bulk of the LFS questionnaire requests information about respondents' activities in a seven day period which ends on a Sunday: this is called a reference week. The majority (about 80%) of interviews are carried out in the week following the reference week, although if this is not possible interviewers are given a further week and two days in which to obtain interviews (known as the hangover period). The hangover period is extended during some weeks leading up to and including Christmas in order to minimise non-contact (in addition, during these periods, face-to-face interviewers use the whole of the hangover period).

Fieldwork documents

In advance of a first interview a letter is sent to every address in the selected sample explaining that the address has been selected and that an interviewer will be calling.

Additionally, in the advance letter, respondents are assured that the information they give will be treated in the strictest confidence and will not be made available to analysts in any form in which individuals, or their households, can be identified. Respondents are also sent a Purpose Leaflet, giving information on summary results and how the LFS data are used (See Annex A for currently used survey documents).

5.2 FIELD MANAGEMENT AND THE LFS SURVEY DESIGN

Avoiding within quarter bias

In any systematic single stage sample of households spread across 13 weeks there is a need to structure the sample so that fieldwork practice does not inadvertently introduce within-quarter bias. One possibility would be to give up the idea of a quarterly sample and simply take un-clustered weekly samples. However, face-to-face interviews for the first wave as well as households needing a face-to-face interview in subsequent waves would amount to a sample of only about 2,400 addresses each week spread over the entire country. The average distance between addresses would then be so great that it would be necessary to train and equip an enormous number of interviewers each of whom would do very few interviews. They would take a very long time to build up useful experience as interviewers, and with such a large number, adequate monitoring and supervision would be difficult. Alternatively with a smaller number of interviewers each would spend most of his/her time travelling between sampled addresses with little or no time to do recalls, leading to heavy non-response bias. Since neither of these options were acceptable to ONS the sample is designed as a series of weekly two stage samples spread over the 13 weeks such that the whole country is covered in the quarter and therefore the quarter as a whole constitutes a single stage sample.

Grouping postcode areas

As noted above, the country is divided up into 208 interview areas each containing an equal number of delivery points working systematically across Great Britain and trying to follow existing regional boundaries as far as possible. Within these 208 areas there is a further sub-division into 412 quotas which are then divided further into 13 "stint" areas by grouping postcode sectors. Again the aim is to create weekly stint areas of equal size in terms of their number of delivery points (though geographical size varies considerably). In order to avoid unnecessary travel problems in the weekly areas, ONS attempted to map out areas so as to make a mountain, lake or other geographical obstacles occur on the border of a stint. Inevitably the stints vary in their make up because some of the larger interviewing areas are either very rural or very urban, but where possible the weekly stints are mapped so that they contain a mixture of urban and rural localities.

5.3 DEPENDENT INTERVIEWING AT RECALL WAVES

The LFS uses dependent interviewing, where answers given at the previous wave are available to interviewers. The use of dependent interviewing has been shown to provide more accurate results than asking the questions from scratch each time. Methodological investigations by the US Bureau of the Census have shown the considerable improvements in the quality of data produced from dependent

interviewing; this technique was recently introduced on their equivalent of the LFS, the Continuous Population Survey (CPS).

Core questions

For most core questions on the LFS the information from the previous wave is rotated into the next quarter. Interviewers must check this information either by asking the question again or checking that the information given in the last wave is still correct.

There are some core questions which have to be asked each quarter without reference to previous answers. These are as follows:

SCHM12	Whether on a work scheme in the reference week
TYPSCH12	Employer of work scheme
YTETJB	Whether had paid work in addition to scheme
WRKING	Whether in paid job
JBAWAY	Whether temporary away from paid job
OWNBUS	Whether doing unpaid work for own business
LEFTW	Whether left last job in reference week
OCCT	Main job in the reference week
HOWGET	How current job was obtained
HOMED	How respondent spent at least one full day at home
ACTWKDY	Days scheduled to work
ILL1PD	Period of sickness
IL1BEF	Period of sickness start day
IL2BEF	First period of sickness start day
ILNXSM	Medical reason
ILCurr	Whether off sick due to work-related illness
TOTAC1	Total actual hours worked in main job
ACTHR	Actual hours worked excluding overtime
ACTPOT	Actual paid overtime in main job
ActUOt	Actual unpaid overtime in main job
YLESS20A	Reason worked fewer hours than usual in reference week
MatLve	Status of maternity leave
YMORE	Reason for working more weekly hours
LssOth	Whether time off was flexi or annualised hours
EVENG	Evening work in last 4 weeks
NIGHT	Night work in the last 4 weeks
SECJOB	Whether had second job in reference week
Y2JOB	Whether had two jobs because of a change of job in reference week
OCCT2	Second job in reference week
ACTHR2	Actual hours in second job including overtime
DIFJOB	Whether looking for a different or additional paid job
LOOK4	Whether looking for any kind of paid work
LIKEWK	Whether would like work
METHMP	Method of looking for work (employees or Government scheme)
METHSE	Method of looking for work (self employment)
METHAL	Method of looking for work (no preference)
START	Whether could start work within the next two weeks
BENFTS	Whether claiming any state benefits/tax credits

UNEMBN	Type of unemployment related benefit claiming
UCREDIT	Reason for claiming Universal Credit
INCSUP	Whether claiming income support in reference week
DISBEN	Type of sickness or disability benefit claimed
HSNGGB	Whether receiving Housing Benefit or Council Tax Benefit
ED4WK	Job related training or education in the last 4 weeks
FUTUR4	Job related training or education in the last 4 weeks
TrHr11	Hours spent on education or training
GROSS99	Gross pay main job
SECGRO	Gross pay second job

5.4 REQUIREMENTS FOR ANSWERS TO QUESTIONS

Whilst every effort is made to obtain answers to all relevant questions from each respondent, it is recognised that there will be some cases when a respondent genuinely does not know the answer to a particular question (particularly in the case of responses by proxy - see below) and cases when a respondent does not wish to give the answer to a particular question. In general ONS would not wish to lose such respondents and a "no answer" or "don't know" will be accepted.

However, there are a number of key questions in the survey, some of which are fundamental in classifying a respondents' economic status, which, if not answered cause that whole record (though not the whole household) to be dropped.

Forced response questions

These 'forced response' questions are currently as follows:

SEX	Sex of respondent
AGE	Age of respondent
HALLRES	Whether living in a hall of residence
MARSTA[13]	Marital status
LIV12W	Whether respondent is living together with someone as a couple
HRPID	Whether accommodation is owned/rented in respondent's name
SCHM12	Whether respondent on a work scheme in the reference week
FUND12	Funding of work schemes
TYPSCH12	Employer of work scheme
WRKING	Whether respondent did any paid work in the reference week
JBAWAY	Whether respondent was away from a paid job in the reference week
OWNBUS	Whether respondent did any unpaid work in the reference week for a business owned by him/herself
RELBUS	Whether respondent did any unpaid work for a business owned by a relative
EVERWK	If ever had paid work
STAT	Whether respondent was working as an employee or self-employed
ILLWK	Had days off work because sick or injured
TOTUS1	Total usual hours worked excluding lunch breaks (no overtime)

---

[13] In the Blaise questionnaire this question is XMARSTA.

USUHR	Usual hours worked excluding overtime
POTHR	Usual hours of paid overtime
UOTHR	Usual hours of unpaid overtime
TOTUS2	Usual hours worked including overtime
TOTAC1	Total actual hours (no overtime)
ACTHR	Actual hours worked excluding overtime
ACTPOT	Actual hours of paid overtime
ACTUOT	Actual hours of unpaid overtime
TOTAC2	Actual hours worked including paid and unpaid overtime
ACTHR2	Actual hours in second job including overtime
UNDHRS	Number of extra hours would like to work
LOOK4	Whether respondent was looking for paid work in the previous 4 weeks
LKYT4	Whether respondent was looking for a place on a Government scheme in the previous 4 weeks
METHMP	Seeking work as an employee
METHSE	Seeking work as self employed
METHAL	Seeking work no preference whether as an employee or self employed
MAINME	Main method of looking for work as an employee
MAINMA	Main method of looking for work as either an employee or self employed
MAINMS	Main method of looking for work as self employed
HPRMB	Prompt to ask health questions
GROSS99	Gross pay before deductions
GRSEXP	Gross pay expected
USUGPAY	What is usual pay received for period covered by last pay
NET99	What is take home pay after deductions
USUNPAY	What would usual amount of net pay for period be
HRRATE	What is basic hourly rate
HRRATE2	Basic hourly rate in second job
SECGRO	Gross pay before deductions
SECEX	Expected gross earnings
SECNET	Net pay after deductions
IREND2	Religious denomination
RELBUP	Religious denomination brought up in

5.5 PROXY INTERVIEWS

Acceptability of proxy responses

The LFS allows interviewers to take answers to questions by proxy if a respondent is unavailable. This is usually from another related adult who is a member of the same household, although there are exceptions to this rule:

    (i)   a person, of the same household, may translate for a non-English speaking relative;

    (ii)   a carer, of the elderly or infirm, although not related, may answer for someone in their care if it can be established that they know the respondent well enough;

(iii)   anyone can respond by proxy with the personal permission of the head of household or spouse.

Frequency of proxy responses

About a third of LFS responses are collected by proxy. This figure includes not only people who were unavailable and on whose behalf a proxy response was made, but also two other groups. The first is those who were unavailable and did not have a proxy response made for them this wave, but did have a proxy response made for them the previous wave, which was brought forward to the current wave. The second group comprises economically inactive individuals aged 70 years or more, for whom proxy responses were given.

Hence the third figure includes all informants for whom proxy data was collected. Information on proxy responses can be obtained using the PRXREL variable. Categories 2 and 3 (defined as 'Spouse/partner proxy' and other proxy' respectively) are combined to give the total number of proxy interviews. Information on proxy responses can also be obtained using the variable IOUTCOME; this identifies the informant's status for a particular variable. However, the main (IOUTCOME=2) category of proxy responses does not include cases where proxy data was imputed from the previous wave, or where proxy responses were given for economically inactive 70+ year olds.

Tables 5.1 and 5.2 show the proxy response rates for Great Britain for different age, sex, ethnic and economic activity categories. The numbers are percentages.

Table 5.1 Proxy response rates (per cent),GB, by age, sex, and ethnicity.

Quarter	Total	Age			Sex		Ethnicity	
		16-17	18-19	20+	Male	Female	White	Non-white
JM17	34	88	77	31	39	30	33	43
AJ17	34	88	78	31	39	30	33	43
JS17	34	89	75	31	39	29	33	43
OD17	34	90	75	31	39	29	33	43
JM18	34	90	76	31	39	29	33	42
AJ18	34	90	78	31	40	29	33	43
JS18	34	90	78	31	39	29	33	43
OD18	34	89	77	31	39	29	33	42
JM19	33	89	77	30	38	29	32	42
AJ19	33	89	77	31	39	29	32	41
JS19	34	89	78	31	39	29	33	42
OD19	33	89	75	30	37	28	32	41
JM20	33	90	78	30	37	28	31	42
AJ20	33	90	79	31	38	29	32	42
JS20	34	91	85	32	40	29	33	43
OD20	35	94	85	33	41	30	34	44
JM21	36	93	88	33	42	31	62	45

AJ21	37	94	90	35	42	31	36	46
JS21	38	96	91	35	43	33	37	47
OD21	38	96	91	35	43	34	37	46
JM22	37	95	92	35	42	33	37	46
AJ22	38	96	92	36	42	34	37	46
JS22	38	97	91	36	42	35	37	46
OD22	37	96	89	35	40	36	36	45
JM23	38	96	91	36	41	36	37	46
AJ23	38	96	93	35	40	36	37	46
JS23	38	97	91	35	40	36	37	46
OD23	37	96	89	35	40	36	36	45
JM24	38	94	87	35	40	35	37	45

Source: Table 7 (Table 9 after JM20) of LFS Performance and Quality Monitoring Reports:
https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeet
ypes/methodologies/labourforcesurveyperformanceandqualitymonitoringreports

Table 5.2 Proxy response rates (per cent) for GB, by employment status.

Quarter	Total	Economic Activity					
		Employees	Self-Employed	Government schemes	Unpaid family workers	ILO Unemployed	Inactive
JM17	34	36	34	36	24	37	31
AJ17	34	35	35	38	28	36	32
JS17	34	36	36	44	21	39	31
OD17	34	35	35	42	25	38	31
JM18	34	35	35	38	21	35	31
AJ18	34	36	36	35	22	36	31
JS18	34	36	36	40	24	38	31
OD18	34	36	37	45	20	39	31
JM19	33	35	35	42	21	36	30
AJ19	33	35	34	43	18	35	31
JS19	34	35	36	41	26	39	30
OD19	33	34	34	46	24	36	29
JM20	33	34	33	35	16	35	30
AJ20	33	34	34	21	19	41	32
JS20	34	35	36	48	23	45	32
OD20	35	36	37	68	19	42	34
JM21	36	37	38	53	28	42	34
AJ21	37	38	41	55	33	45	35
JS21	38	39	41	54	32	46	35
OD21	38	39	41	57	27	45	36
JM22	37	39	40	67	24	42	35

AJ22	38	39	40	51	28	43	36
JS22	38	40	39	51	22	46	36
OD22	39	40	39	33	27	48	36
JM23	38	41	38	35	27	46	35
AJ23	38	40	39	40	38	43	36
JS23	38	40	40	55	31	48	35
OD23	37	40	40	36	37	51	34
JM24	38	40	41	37	25	45	34

Source: Table 7 of LFS Performance and Quality Monitoring Reports:
https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/met
hodologies/labourforcesurveyperformanceandqualitymonitoringreports

Further information about proxies, especially quality of data, can be found in the
report 'A study of proxy response in the Labour Market Survey' which was an article
written by Fiona Dawe and Ian Knight and was published in the Survey Methodology
Bulletin (No.40), January 1997.
http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/o
ns/guide-method/method-quality/survey-methodology-bulletin/smb-40/index.html

5.6 COMPUTER ASSISTED INTERVIEWING (CAI)

The LFS interviews are carried out by face-to-face and telephone interviewers using
laptop computers. The questionnaire and edit instrument that they use is identical in
both modes. It is produced using the BLAISE CAI software package, which was
created by Statistics Netherlands. SSD has designed a computer system which
takes the output from BLAISE and uses it to create derived variables, to weight up
population estimates and other processes leading to the production of data files for
customers. The system also rotates the data for use at the next wave of interviewing.

Advantages and disadvantages of CAI
The advantages of CAI for the LFS over a paper-and-pencil based system are lower
costs, improved speed from fieldwork to analysis and better quality data.
Disadvantages such as the initial cost of equipment are outweighed by the
advantages for an ongoing, regular survey such as the LFS. Improved quality arises
from the completion of editing in the interview, where inconsistencies can be
checked with the respondent and the use of automatic routing to ensure that
respondents are asked all of the relevant questions. Data capture and editing in the
interview, and electronic transmission of the data, are the main contributions to
improved speed of delivery results.

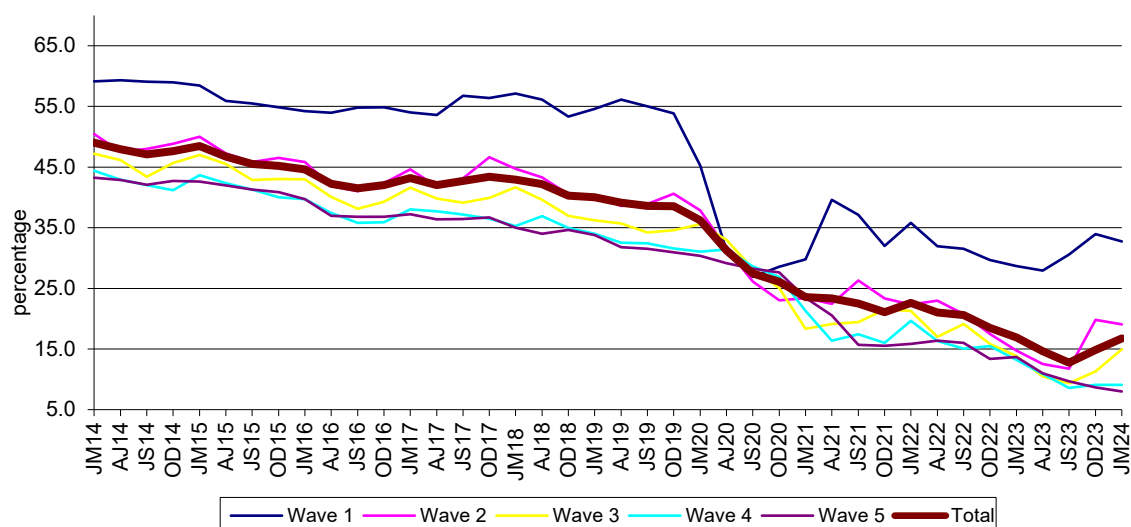## 5.7 RESPONSE RATES

Panel survey non-response

As the LFS is a panel survey, the calculation of response rates should take the panel design into account. Households may refuse further participation at any of the five quarterly visits they are due to receive.

Households which refuse further participation are not revisited at the next quarter but they remain part of the eligible sample. The response rate for households comprises the ratio of the number of households responding at the current wave to the sum of the number of eligible households found at the same wave at the sampled addresses, plus any households which have refused outright to participate at a previous wave. Outright refusals (as distinct from circumstantial refusals and non-contacts, which are revisited at the next wave) may occur either when the interviewer calls or asks for permission to recall in three months.

The simple model above does not take account of such situations as net addition or subtraction of eligible households at sampled addresses at waves after wave 1, for example by a net increase or decrease in occupation of household spaces. However, the definition of household spaces is fluid and does not provide a firm basis for response rate calculations. LFS response rates which take the panel design into account are based on the simple model.

Figure 5.1 plots wave specific LFS response rates from Q1 2014 to Q1 2024. This includes both face-to-face and telephone interviewing and applies to cases in Great Britain only. Table 5.3 presents the data underlying this figure.

Figure 5.1: LFS Quarterly Survey, wave specific response rates, Great Britain, excluding imputed households JM14 to JM24

Table 5.3: LFS Quarterly Survey, wave specific response rates, Great Britain, excluding imputed cases OD09 to JM24[14]

Period	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Total
						per cent
OD09	62.4	57.5	53.1	50.3	50.1	54.7
JM10	64.6	58.1	55.8	51.3	50.2	55.9
AJ10	64.5	56.9	53.2	50.6	48.8	54.8
JS10	63.9	54.9	49.8	46.3	45.7	52.2
OD10	63.5	52.9	49.5	45.0	44.4	51.2
JM11	61.5	51.9	48.2	44.9	42.7	50.0
AJ11	60.9	48.7	47.3	42.9	42.4	48.6
JS11	62.8	51.7	47.4	46.0	43.9	50.5
OD11	62.6	53.7	49.8	45.4	48.0	51.5
JM12	62.1	52.5	50.4	46.8	44.1	51.3
AJ12	60.9	51.6	48.9	46.5	45.0	50.7
JS12	58.9	51.0	48.5	45.7	45.0	49.9
OD12	56.9	49.5	47.4	44.0	43.7	48.4
JM13	57.3	50.8	48.1	45.0	44.4	49.3
AJ13	56.5	47.7	46.7	44.3	43.5	47.9
JS13	58.9	50.2	46.6	45.1	45.2	49.3
OD13	59.4	50.4	46.5	43.4	43.9	48.8
JM14	59.1	50.5	47.2	44.4	43.2	49.0
AJ14	59.3	47.5	46.1	42.9	42.9	47.9
JS14	59.1	48.0	43.4	42.0	42.1	47.1
OD14	58.9	48.8	45.7	41.2	42.7	47.6
JM15	58.4	50.0	47.0	43.7	42.6	48.5
AJ15	55.9	47.2	45.5	42.3	42.0	46.7
JS15	55.5	45.8	42.9	41.3	41.3	45.5
OD15	54.8	46.5	43.1	40.0	40.9	45.2
JM16	54.2	45.8	43.0	39.7	39.7	44.6
AJ16	54.0	41.8	40.1	37.4	36.9	42.2
JS16	54.8	41.0	38.1	35.8	36.8	41.5
OD16	54.9	42.4	39.3	35.9	36.8	42.0
JM17	54.0	44.6	41.6	38.0	37.2	43.2
AJ17	53.6	41.8	39.8	37.7	36.4	42.0
JS17	56.7	43.1	39.1	37.2	36.5	42.7
OD17	56.4	46.6	39.9	36.5	36.7	43.4
JM18	57.1	44.7	41.6	35.2	35.0	42.9
AJ18	56.1	43.3	39.6	36.9	34.0	42.2

[14] Wave specific response rates for periods before OD09 can be found in an earlier version of this user guide.

JS18	53.9	40.0	36.6	34.1	34.4	40.0
OD18	53.3	40.5	36.9	35.0	34.6	40.3
JM19	54.6	40.3	36.2	34.0	33.8	40.0
AJ19	56.1	38.6	35.7	32.5	31.8	39.1
JS19	55.0	38.9	34.2	32.4	31.5	38.6
OD19	53.8	40.6	34.6	31.6	30.9	38.5
JM20	29.8	23.4	18.3	21.3	23.4	23.6
AJ20	31.2	31.6	33.0	31.4	29.1	31.3
JS20	26.9	26.1	28.2	28.6	28.3	27.5
OD20	28.6	23.0	25.0	26.9	27.7	26.1
JM21	29.8	23.4	18.3	21.3	23.4	23.6
AJ21	39.6	22.5	19.1	16.4	20.5	23.3
JS21	37.1	26.3	19.4	17.4	15.7	22.5
OD21	32.0	23.3	21.5	16.0	15.5	21.1
JM22	35.8	22.4	21.3	19.6	15.8	22.6
AJ22	31.9	22.9	17.0	16.3	16.4	21.0
JS22	31.5	20.8	19.1	15.1	16.0	20.6
OD22	29.7	17.5	15.8	15.5	13.4	18.5
JM23	28.7	14.7	13.8	13.2	13.7	16.9
AJ23	27.9	12.5	10.5	10.9	11.0	14.6
JS23	30.5	11.7	9.3	8.6	9.7	12.7
OD23	34.0	19.8	11.3	9.1	8.7	14.8
JM24	32.7	19.7	14.9	9.1	8.0	16.7

Source: Table 2 of LFS Performance and Quality Monitoring Reports

Figure 5.2 and Table 5.4 show the number of responding households while Figure 5.3 and Table 5.5 show the number of responding persons. Both figures and tables separately identify imputed cases from personal interviews.

Figure 5.2: LFS Quarterly Survey, achieved number of household interviews, GB and UK, JM14 to JM24
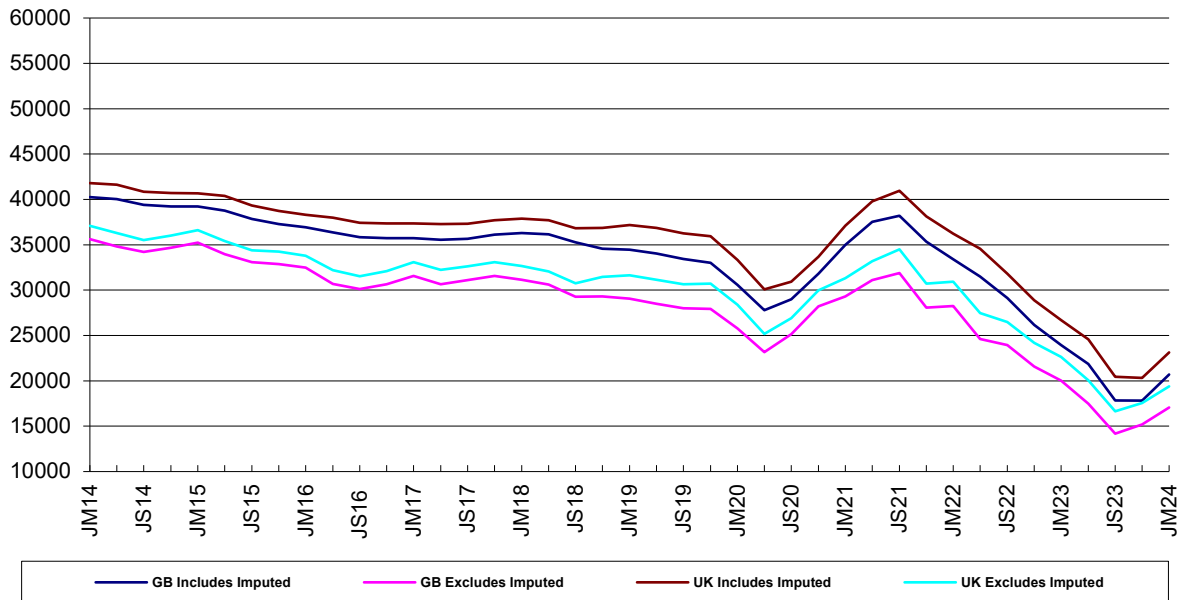
Source: Figure 1 of LFS Performance and Quality Monitoring Reports

Figure 5.3: LFS Quarterly Survey, achieved number of person interviews, GB and UK, JM14 to JM24
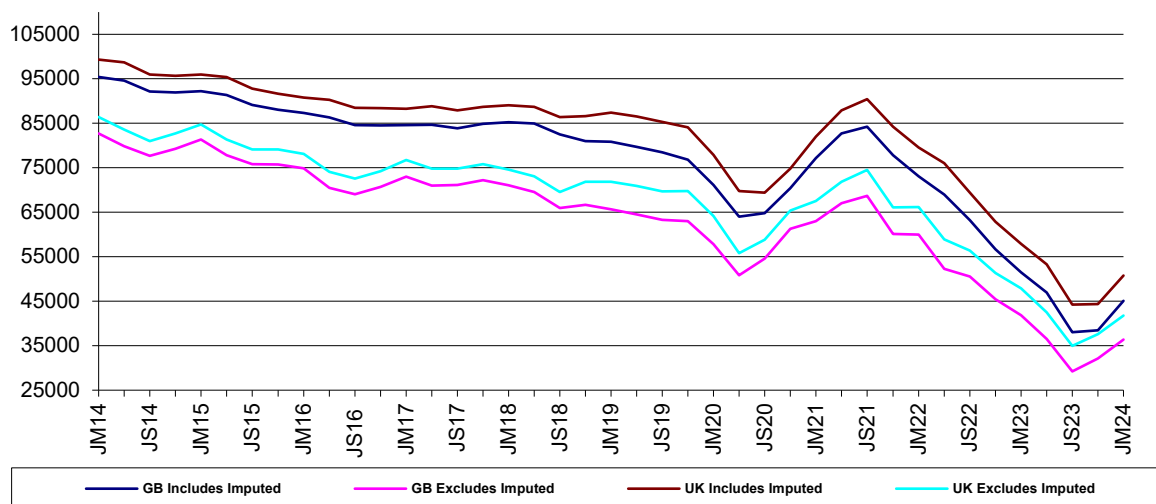
Source: Figure 2 of LFS Performance and Quality Monitoring Reports

Table 5.4: LFS Quarterly Survey, achieved number of household interviews, GB and UK, OD09[15] to JM24

Period	Total		Total	
	GB Includes imputed	GB Excludes Imputed	UK Includes imputed	UK Excludes Imputed
OD09	46,809	42,530	48,584	44,223
JM10	46,851	43,414	48,618	45,109
AJ10	46,934	42,631	48,803	44,433
JS10	42,582	38,404	44,262	40,012
OD10	42,165	37,413	43,746	38,915
JM11	41,691	36,277	43,319	37,815
AJ11	41,267	35,245	42,899	36,803
JS11	41,652	36,615	43,264	38,188
OD11	42,047	37,384	43,635	38,900
JM12	42,216	37,476	43,795	38,976
AJ12	41,939	36,873	43,645	38,480
JS12	41,033	36,219	42,655	37,767
OD12	40,360	35,129	41,999	36,685
JM13	40,279	35,734	41,875	37,221
AJ13	39,885	34,695	41,527	36,186
JS13	39,940	35,720	41,509	37,186
OD13	40,244	35,452	41,850	36,936
JM14	40,254	35,624	41,800	37,090
AJ14	40,053	34,825	41,633	36,291
JS14	39,385	34,227	40,865	35,521
OD14	39,241	34,667	40,711	36,016
JM15	39,216	35,250	40,678	36,608
AJ15	38,770	33,973	40,373	35,425
JS15	37,852	33,078	39,343	34,405
OD15	37,284	32,889	38,741	34,258
JM16	36,922	32,495	38,319	33,797
AJ16	36,381	30,702	37,977	32,196
JS16	35,839	30,105	37,409	31,532
OD16	35,746	30,646	37,342	32,114
JM17	35,738	31,560	37,360	33,102
AJ17	35,572	30,634	37,287	32,231
JS17	35,659	31,099	37,316	32,634
OD17	36,111	31,555	37,697	33,072
JM18	36,312	31,156	37,896	32,652
AJ18	36,165	30,616	37,705	32,065
JS18	35,259	29,263	36,832	30,739
OD18	34,582	29,310	36,869	31,471
JM19	34,473	29,065	37,167	31,620
AJ19	34,021	28,482	36,846	31,137
JS19	33,444	27,991	36,269	30,654
OD19	33,002	27,944	35,955	30,717
JM20	30,583	25,777	33,329	28,385
AJ20	27,798	23,154	30,093	25,167
JS20	28,992	25,136	30,923	26,910

[15] The achieved number of household interviews for periods before OD09 can be found in an earlier version of this user guide.

OD20	31,808	28,223	33,685	29,993
JM21	34,987	29,324	37,098	31,335
AJ21	37,529	31,097	39,799	33,207
JS21	38,208	31,893	40,940	34,484
OD21	35,335	28,084	38,136	30,735
JM22	33,407	28,235	36,236	30,929
AJ22	31,502	24,607	34,552	27,491
JS22	29,129	23,957	31,820	26,500
OD22	26,173	24,586	28,879	24,177
JM23	23,957	20,015	26,712	22,655
AJ23	21,855	17,470	24,572	20,060
JS23	17,824	14,180	20,453	16,645
OD23	17,822	15,182	20,324	17,544
JM24	20,685	17,046	23,143	19,378

Source: Table 1 of LFS Performance and Quality Monitoring Reports.

Table 5.5: LFS Quarterly Survey, achieved number of person interviews, GB and UK, OD09[16] to JM24

	Total		Total	
	GB Includes imputed	GB Excludes Imputed	UK Includes imputed	UK Excludes Imputed
OD09	106,112	94,056	110,511	98,237
JM10	106,457	96,889	110,817	101,064
AJ10	106,531	94,061	111,177	98,529
JS10	99,568	87,919	103,943	92,107
OD10	98,801	85,731	102,842	89,560
JM11	97,432	82,763	101,645	86,727
AJ11	96,032	80,069	100,286	84,113
JS11	96,343	82,939	100,543	87,048
OD11	97,950	85,416	102,059	89,328
JM12	98,487	85,630	102,532	89,468
AJ12	98,105	84,124	102,427	88,206
JS12	95,773	82,917	99,901	86,845
OD12	94,378	80,406	98,513	84,350
JM13	94,494	82,219	98,531	85,984
AJ13	93,121	78,994	97,274	82,740
JS13	93,432	82,223	97,380	85,897
OD13	94,826	81,698	98,878	85,435
JM14	95,411	82,706	99,315	86,398
AJ14	94,621	79,818	98,668	83,587
JS14	92,143	77,698	95,950	80,989
OD14	91,953	79,298	95,704	82,717
JM15	92,233	81,356	95,941	84,757
AJ15	91,361	77,816	95,359	81,364
JS15	89,096	75,835	92,784	79,105
OD15	88,022	75,766	91,618	79,139

---

[16] The achieved number of person interviews for periods before OD09 can be found in an earlier version of this user guide.

JM16	87,299	74,902	90,787	78,141
AJ16	86,349	70,456	90,283	74,094
JS16	84,572	69,076	88,465	72,554
OD16	84,518	70,707	88,406	74,225
JM17	84,556	72,976	88,528	76,735
AJ17	84,661	70,956	88,801	74,785
JS17	83,838	71,101	87,899	74,832
OD17	87,875	72,191	88,726	75,873
JM18	85,203	71,051	89,039	74,619
AJ18	84,937	69,521	88,705	73,041
JS18	82,469	65,956	86,377	69,565
OD18	81,026	66,672	86,585	71,870
JM19	80,849	65,688	87,417	71,844
AJ19	79,665	64,520	86,548	70,943
JS19	78,441	63,256	85,342	69,679
OD19	76,794	63,005	84,062	69,783
JM20	71,105	57,783	77,903	64,158
AJ20	64,040	50,871	69,733	55,837
JS20	64,765	54,587	69,425	58,846
OD20	70,431	61,258	74,832	65,362
JM21	77,192	62,986	82,015	67,549
AJ21	82,715	67,049	87,904	71,817
JS21	84,223	68,712	90,453	74,537
OD21	77,810	60,088	84,205	66,079
JM22	73,053	59,983	79,561	66,145
AJ22	68,943	52,277	75,992	58,900
JS22	63,211	50,538	69,437	56.362
OD22	56,642	45,454	62,862	51,374
JM23	51,513	41,842	57,876	47,910
AJ23	46,964	36,526	53,287	42,512
JS23	38,034	29,220	44,238	34,965
OD23	38,467	32,136	44,338	37,628
JM24	45,057	36,378	50,783	41,752

Source: Table 1 of LFS Performance and Quality Monitoring Reports.

The decline in response rates has been explored in the National Statistics Quality Review :
http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-reviews/list-of-current-national-statistics-quality-reviews/nsqr-series--2--report-no--1/index.html

5.8 QUALITY CONTROL OF FIELDWORK

The LFS is a high-quality product. This has been achieved over a period of time with methodological research and continued improvements to both fieldwork and management practices. For example, in order to minimise non-response, interviewers call back at non-contactable addresses a minimum of three times, across different timeslots. All interviewers have been trained in Achieving Cooperation Training (ACT).

Supervision and training are an important determinant of quality control too; the work of all interviewers is regularly monitored in respect of interviewing technique, dealing with the public, response rates, work efficiency, calling patterns (field interviewers only) and the quality of completed work, including the accuracy of coding. Interviewers are provided with both verbal and written feedback on their performance. Where a weakness in performance is identified, additional training and monitoring is carried out.

5.9 NON-ENGLISH SPEAKING RESPONDENTS

Measures to meet the Welsh Language Act

Since 1985, all sample addresses in Wales are sent advance letters in both Welsh and English. A Welsh translation of the Purpose leaflet is also sent. Where a respondent requests that the interview be conducted in Welsh, arrangements can be made to transfer the household to a Welsh speaking interviewer. However, such requests are rare.

Measures to gain response from non-English speakers

All face-to-face interviewers are issued with a language identification card, containing a message written in the 22 main foreign languages spoken in Great Britain[17]. The card is used to identify a time when an English-speaking family member or friend can be contacted to explain the survey's purpose. Where there is no English speaker available, the card also enables interviewers to identify the language spoken so that the interviewer can arrange an interpreter.

---

[17] These languages include: Arabic, Bengali, Chinese (traditional), Chinese (simplified), Czech, Greek, Gujarati, Hindi, Lithuanian, Polish, Portuguese, Punjabi, Romanian, Russian, Serbian, Slovak, Somali, Tamil, Turkish, Urdu, Welsh, Yoruba.

SECTION 6 - CODING AND PROCESSING THE DATA

6.1 CODING

All coding of data, where required, is carried out by interviewers; some of this is performed using Computer Assisted Coding (CAC) during the interview while the rest is carried out after the interview.

The following questions currently require coding and the method currently used is explained:

Variable	Description	Coding method:	
		BY CAC DURING INTERVIEW	by interviewer after interview
CRYO7/CRYSPEC	Country of Birth	✓	
NATO7/NATSPEC	Nationality	✓	
NATIDCOD/NATIDO	National Identity	✓	
ETHO2/ETHDES	Ethnicity	✓	
RELOCOD/RELOTH	Religion	✓	
M3CRYO/M3CRYSPEC	Country of residence 3 months ago (if outside UK)	✓	
M3RESC/M3AREA/M3CTY	Place of residence 3 months ago	✓	
OYCRYO/OYCRYSPEC	Country of residence 1 year ago (if outside UK)	✓	
OYRESC/OYAREA/OYCTY	Place of residence 1 year ago	✓	
WKPL99/WKARBC/ WKTOWN/WKCTY	Workplace of main job	✓	
WKPL299/WKTOW2/WKCTY2	Workplace in second job	✓	
SMSOC10/SRCHSMSOC/SMOCCT/ SMOCCD	Occupation of main wage earner (social mobility)	✓	
SUBCODE/SUBJQ	Area of study	✓	
SNGDEGN/FDSNGDEG/SNGHD/S UBJCTN/FDSUBJ/HDSUBJCT	Subject Area	✓	
UGINST/PGINST/FDINST/HDINST	Institution studied at	✓	
CURCODE/CURSUB	Subject of qualification	✓	

TSUB4COD/TSUBJ4WK	Subject of tuition	✔	
TIMECODE/TIMEDAYS	Time after accident until returned to work	✔	
INDD/INDT	Industry in main job		✔
OCCT/OCCD	Occupation in main job		✔
RDINDD/RDINDT	Industry before redundancy		✔
RDOCCT/RDOCCD	Occupation before redundancy		✔
INDD2/INDT2	Industry in second job		✔
OCCT2/OCCD2	Occupation in second job		✔
OYINDD/OYINDT	Industry in job 1 year ago		✔
OYOCCT/OYOCCD	Occupation in job 1 year ago		✔
APPD/APPT	Apprenticeship (continuing)		✔
APPIND/APPINT	Apprenticeship industry		✔

6.2 DERIVED VARIABLES

In order to analyse LFS data, a number of derived variables (DVs) are specified. The survey cannot provide all of the information that users want by asking a simple question, there DVs are created by combining the answers to two or more questions from the questionnaire. These do not vary significantly from year to year, although only those DVs which relate wholly to core data are created each quarter. DVs relating to non-core data are only created in the quarter(s) those non-core questions are included. All DVs are specified for the UK as they are created after the NI data are merged with GB data. Volume 4 contains the latest set of flow diagrams used to specify the current DVs, which can be split into four groups:

Person: DVs created for each individual record on the database;

Family: DVs created for each family on the database; on a flat file each family member would carry the same value for each family based DV;

Household: DVs created for each household on the database; on a flat file each household member would carry the same value for each household based DV;

Eurostat: DVs were created for each record specifically to meet Eurostat requirements. These are not currently available to external customers: they are included here for completeness.

6.3 DATA CHECKING

Whilst some checking is performed in-the-field by the BLAISE survey instrument, other checks are carried out once the data have been received back from interviewers in the field or from the telephone unit.

The principles of the checks are to ensure that the data have no duplication of records etc, that the data have the correct household structure, in terms of persons in the household, and that certain key variables have valid values. These checks are important in maintaining the quality of the data. Examples include:

- checking that families have been correctly assigned within households by referring to the relationship grid
- checking that responses from a previous wave are consistent with the current wave on a number of key variables (eg. AGE, SEX ...). These are known as 'imputation checks'
- country checks to ensure geographies are consistent across variables
- checks to ensure each household has a household reference person (HRP)
- checks to ensure that there is a record for every member of a household – known as the 'full house check'
- checks to ensure that the household type (HHTYPE6) matches the household composition as defined in the relationship grid
- checks on ages for outliers (e.g. very old, old workers etc)
- cross checks between variables indicating activity status (e.g. INECAC05 and STATR & INECAC05 and ILODEFR)

SECTION 7- NON-SAMPLING ERRORS

The following section is drawn from the report 'Measuring and Improving Data Quality' by Vera Ruddock published as part of the GSS Methodology Series (no. 14)[18].

The provision of accurate, timely data which meets the needs of users at minimal cost is at the heart of government statistics. There are two components to accuracy: sampling errors which occur when data from a sample is used to make inferences about the whole population (see section 8) and the so called 'non-sampling errors' which affect data from sample surveys, as well as administrative and census data. Non-sampling errors should not be viewed as mistakes, rather they are the result of conscious decisions to produce timely, accurate data at minimum cost.

Measuring non-sampling error is much more difficult than measuring sampling error because in many cases the reasons for the non-sampling error are not known, whereas sampling error is a direct result of the survey design and is under the control of the researcher. In some cases it may not be possible to measure non-sampling error or to only give an indication of its possible effect on the survey estimates. Non-sampling errors can also be very expensive to measure.

The rest of this section examines how accuracy of survey estimates can be measured and describes the different types of non-sampling error and their occurrence on the LFS.

7.1 MEASURING ACCURACY

Users of statistics commonly ask the question 'Is the estimate accurate?' The answer to this question influences the value the user attaches to the estimate, and the potential for the estimate to change the user's beliefs about a given subject. Accuracy is one concept, which defines the quality of a survey estimate. Accuracy reflects the difference between the survey estimate and the population parameter being estimated.

The question 'Is the estimate accurate?' is only the first part of the underlying question 'Is the estimate accurate enough for the purpose I want to use it for?' Discussions of the accuracy of estimates must therefore reflect the context in which they are to be used, but this assessment of the suitability of estimates for addressing specific issues requires some measure of the quality of the data. Accuracy is not usually reported; instead the error in an estimate is described by the bias and variance in that estimate, the two components of the total survey error.

---

[18] http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/method-quality/specific/gss-methodology-series/index.html

Total Survey Error is the inverse of accuracy. A statistic with low accuracy will have high total survey error. Total survey error is measured by the mean square error, which is defined as the sum of all biases and variances:

i.e. MSE = variance + bias^2

The biases and variances may be due to sampling error, non-sampling error or both. Reported mean squared errors for survey estimates commonly only include the bias and variance attributable to sampling error. However non-sampling error can lead to biased estimates, for example in surveys people may systematically under report their consumption of alcohol. Similarly slight differences in the way respondents react to different interviewers may lead them to give different answers to different interviewers resulting in interviewer variance.

7.1.1 Bias

Sources of bias can be classified into errors of non-observation and errors of observation.
Errors of non-observation include:
- coverage error. If the register or sample frame used to select the sample does not represent all the target population then the resulting sample estimate may be biased. For example, a random sample of people who are in the telephone directory will exclude both those who:
- have no telephone and those who are ex-directory. If the value of a variable being measured in the survey is different for people who are and are not in the telephone directory, then the survey estimate will be biased (see Sampling Frames in section 3);
- non-response. If the people who do not respond to surveys are different from responders, then estimates from the achieved sample may be biased estimates of population values - this can be corrected to some extent by weighting the sample.

   Errors of observation include:
- social desirability effects: an unwillingness of respondents to admit to socially undesirable behaviour. This is most obvious in surveys of sexual experience when men commonly over-report and women under-report the number of sexual partners they have had in their lifetime.

7.1.2 Variance

An estimate of the variance of a statistic is based on the variability within the sample, which arises because achieved values differ over the units (e.g. sampled person, interviewers used, questions asked) that are the sources of the error.

Sources of variance include:
- Sampling variance:
  In sample surveys only a proportion of the population has been sampled. The sampling variance reflects the fact that the estimate may have been different if a different sample had been selected.
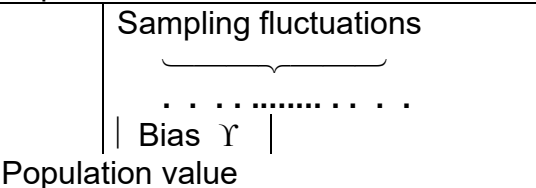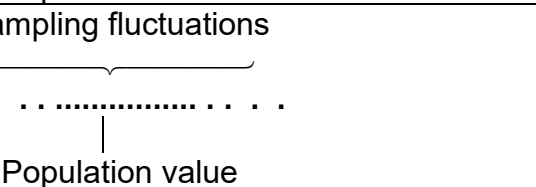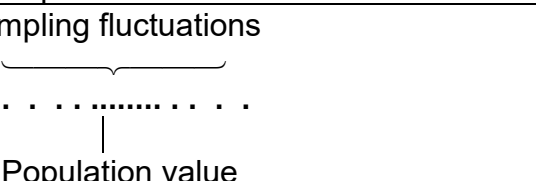- Non-sampling variance:
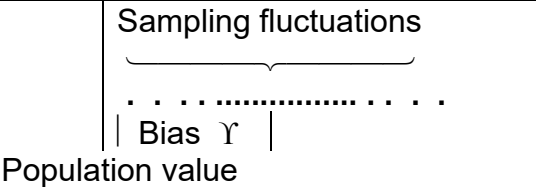
There are a variety of sources of non-sampling variance. For example, differences between interviewers may consciously or unconsciously cause variation in the answers given by respondents; this is known as interviewer variance since it is due to differences between interviewers in achieved responses.

In the course of designing a survey many decisions are taken which may influence the relative size of different sources of error. These may reduce the bias in an estimate at the expense of an increase in the variance of the estimate. Alternatively, survey designs which minimise one source of survey error may lead to an increase in another source of error. For instance, a common trade-off is the issue of whether to allow proxy responses in household surveys.

In the LFS adult members of a household are allowed to answer questions on behalf of absent members of the household. This minimises the extent of missing data (item non-response), but the quality of the data from proxy respondents is not always as high as data from the actual intended respondent so respondent error is increased. In contrast to the LFS, proxy responses are not accepted on the Living Costs and Food Survey (LCF), because of the very detailed nature of the survey, the result being a much lower response rate.

In some cases the use of proxies may introduce respondent bias into the results if they consistently underestimate variables such as household income, but in others some proxies will underestimate and others overestimate the true value leading to an increase in respondent variance. For more discussion of proxy responses on the LFS, see section 5.5.

The following diagram (from Moses) illustrates how bias and precision relate to distinct aspects of sampling procedure.

A Large bias, low precision		Sampling fluctuations . . . . | Bias ϒ | Population value
B Large bias, higher precision		Sampling fluctuations . . . . | Bias ϒ | Population value
C No bias, low precision	Sampling fluctuations . . . . Population value	
D No bias, higher precision	Sampling fluctuations . . . . Population value	

The dots in the diagram represent estimates of the population value derived from repeated application of the given survey procedures i.e. they represent the sampling distribution of the estimates and their mean is the expected value. The population value is what the survey is trying to estimate. The terms 'large', 'low' etc are, of course, relative.

It can be taken that (D), which is unbiased and relatively precise, is the ideal, whereas (A) is to be avoided. In practice the choice is not as simple as this, and there are circumstances in which a sample designer might be prepared to tolerate some bias if precision could markedly be increased.

7.2 TYPES OF NON-SAMPLING ERROR

Non-sampling errors - bias and variance - can be classified into three broad categories:
a) errors of non-observation, which can be broken down into:
- coverage error
- non-response error

b) measurement errors, of which there is:
- interviewer error
- respondent error
- instrument (or questionnaire) error
- mode error

c) processing errors, consisting of:
- systems error
- data handling error

Each of these is described below, in relation to the LFS.

7.3 ERRORS OF NON-OBSERVATION

7.3.1 Coverage error

Coverage error is the error which arises because some units are either excluded or duplicated on the sampling frame used to identify members of the population of interest.

A sampling frame has 3 elements:
- a list representing all elements in the target population;
- further characteristics of these elements (auxiliary information);
- the probability of selecting each element on the frame.

The coverage ratio is the proportion of the target population included on the sampling frame. It gives an indication of the level of possible under-coverage, but does not measure the impact of under-coverage on survey estimates. Biased estimates can be caused by under coverage and duplicate listings while increased

variance of estimates can be caused by inclusion of non-population elements in the list and errors in auxiliary information.

Coverage bias and variance can be measured by comparing data on the sampling frame with external data and by using special data collection procedures incorporated into the survey. Coverage error is minimised by using accurate up-to-date frames. Out of date lists can have the following impact on data quality:

- bias in survey estimates if new elements differing from elements already on the list have not been added to the list;
- increase in the variance of estimates if auxiliary information used for stratification or estimation is inaccurate, or it is discovered during a survey that sampled elements should not be on the list;
- reduction in survey response rate if elements are untraceable and it is not possible to ascertain that the elements are either old elements which should not be on the sampling frame or elements with inaccurate addresses which cannot therefore be traced.

For more detail on the LFS sampling frame and the way it is kept up-to-date, see section 3 of this volume.

7.3.2 Non-Response error

There are two types of non-response error:

- Unit non-response: failure to obtain any of the substantive measurements from the sampled unit (the unit response rate is the proportion of the sampled population responding to a survey);
- Item non-response: failure to obtain specific items of information from an otherwise responding unit.

Non-response bias in an estimate has two components:

- the proportion of the sample responding to the particular question;
- the difference between the true answer to a question in respondents and non-respondents.

Even if the response rate is high, large differences in the true answer to a question in respondents and non-respondents may lead to substantial non-response bias. Non-response can reduce the precision of survey estimates - this can be pre-empted at the design stage by increasing the size of the survey sample.

Unit non-response
There are four sources of information about non-respondents which can be used to examine the existence of unit non-response bias:

- information on the sampling frame;
- census records for responding and non-responding units which can be matched to the sampling frame;
- information collected by interviewers in a follow up survey of non-respondents;
- in panel surveys, information collected from respondents in earlier waves of the survey.

Strategies for minimising unit non-response include:

- interviewer training to reduce the number of refusals - interviewer training on the LFS is rigorous and all interviewers work solely on the LFS.
- encouraging interviewers to call on weekday evenings and at weekends - the timings of the LFS face-to-face and telephone interviews are managed in order to maximise the chances of gaining a response from a household, so much of the interviewing is done in the evenings.
- sending an informative well designed advance letter in interview surveys – households chosen for the LFS are sent a letter before their first interview which explains the background to the LFS, that the survey is voluntary and that responses will be treated as confidential, and gives a rough idea of when the interviewer is likely to call.
- reducing the burden on the potential respondent - the length of the questionnaire is reviewed regularly in order to keep the interview length down.
- offering incentives to respondents - while respondents to the LCF are given a monetary reward for completing a diary of their spending, no incentives or rewards are offered to LFS respondents.
- sending follow up reminders for postal questionnaires and making repeat calls in telephone/face-to-face questionnaires. Face-to-face and telephone interviewers will make a number of attempts to contact a household before it is treated as non-response.

For more information on LFS interviewing, see section 5. Weighting is also used on the LFS (see section 10) to compensate for unit non-response. The complicated population weighting allocates a weight to each individual, ensuring that the respondents are representative of the population as a whole, in terms of age, sex and region of residence. It also converts the sample estimates into estimates expressed in terms of the population.

Item non-response
An indication of the level of item non-response bias can be gained by comparing the characteristics of people responding and not responding to a particular question. Strategies for minimising item non-response include:
- clear question design;
- computer assisted modes of administering interviews to reduce routing errors and identify possibly erroneous data in the course of the interview - all LFS data is collected by laptop or PC, allowing a number of data checks to take place during the interview.

On the LFS, imputation (see section 12) is used to estimate missing items on a questionnaire so that the potential bias in estimates due to item non-response may be reduced.

7.4 MEASUREMENT ERROR

There are four types of measurement error:
- interviewer error arising from both conscious and unconscious differences in the way interviewers administer a survey, and also from the reactions of respondents to different types of interviewers;
- respondent error arising from the inability or unwillingness of a respondent to produce a correct answer;
- instrument error which reflects the effect of question wording, response categories and form design on responses; and
- mode error which describes the effect of different methods of administering a questionnaire on the recorded responses.

Measurement bias can only be accurately measured in record check studies where the true value of a response is matched to the survey response. An indicator of measurement bias can be obtained from split sample studies where one component of the survey design is varied across subgroups of the sample. Different subgroups may:
- receive different questionnaires to investigate instrument bias; or
- have their interview administered in different ways, for example some may receive a face-to-face interview and others may fill in a self-completion questionnaire to investigate mode bias.
- cognitive testing methods which ask respondents to questionnaires why they gave certain answers and attempt to understand the process leading to a response may be used to study respondent and instrument bias.
- measurement variance is important in interviewer surveys. High interviewer variance can have a large effect on the precision of survey estimates.

The different types of measurement error can be minimised using a variety of methods:
- interviewer error is minimised by thorough ongoing interviewer training and the use of small interviewer quotas to reduce the influence of interviewer variance on the precision of survey estimates;
- respondent and instrument error are minimised by careful question testing – new questions for the LFS are generally tested twice and feedback from the interviewers taken into account before the questions become part of the survey.
- mode error is minimised by using appropriate methods to collect data on sensitive questions - it is hoped that response bias on LFS earnings questions can be evaluated and perhaps reduced by asking respondents to check documentary evidence such as a payslip and recording whether such evidence was provided. There are also small but noticeable differences in the information collected by face-to-face interviewers and by telephone interviewers. Although some of the difference can be explained by respondents getting used to the interviewing process with each successive

quarter's questioning, some of the difference is also due to the mode effect and it is difficult to disentangle the two causes. Estimates of employment are about 1 per cent lower, on average, in first interviews (face-to-face) than in subsequent interviews (telephone). However, as the survey design has not changed in recent years the estimates are consistent over time, and therefore estimates of change are unaffected by these effects.

7.5 PROCESSING ERROR

There are two types of processing error: systems error and data handling error. Systems errors are errors in the specification or implementation of systems needed to carry out surveys and process results; system errors on the LFS can creep in when derived variables are specified and/or amended. Data handling errors are errors in the processing of survey data.

There are various sources of data handling error:
- *Data capture*
    Information recorded on a paper questionnaire may be inaccurately converted to a format which can be interpreted by a computer. On the LFS, data capture is automatically incorporated into computer assisted interviewing modes of data collection, but interviews themselves may mis-key answers. This type of error on the LFS is minimised by using mainly computer-assisted data capture with inbuilt checks.
- *Data transmission*
    Electronic data on interviews may be lost in transit between the field and the head office but this can be minimised by using an effective case management system to track the progress of individual packets of data.
- *Editing*
    Errors may be introduced when raw survey data is transformed into a dataset which can be used for producing estimates. These errors can be minimised by:
    - incorporating survey edits into computer assisted interviews so that the respondent can be asked about suspect responses - the method used on the LFS;
    - involving subject matter specialists so that the edits are appropriate for the data;
    - testing program code used in editing.
- *Coding*
    Coding is the transformation of textual open-ended responses to survey questions into categories to be used in data analysis. Coding systems may be manual, computer assisted - where the computer suggests a list of possible codes to the human coder, or computer automated. The last two of these methods are used on the LFS, particularly for industry and occupation coding.

    Individual coders may unconsciously show preferences for particular codes. The impact of these individual biases in the codes allocated by coders on survey estimates may cancel out, however although the survey estimate may not be biased, the variance of the estimate may be

increased. If the individual biases do not cancel out then the coding error will introduce bias into the survey estimate. These types of errors can be minimised by effective training of coders in using the coding system.

- *Weighting and imputation*
  The use of inappropriate methods of weighting and imputation may introduce errors into survey estimates. See section 10 for more detail on LFS weighting and section 12 on imputation.

SECTION 8 - SAMPLING ERRORS AND CONFIDENCE INTERVALS

Surveys are prone to errors arising from a number of sources and processes. Frequently a distinction is drawn between *non-sampling errors* and *sampling errors*. Sampling errors result from the fact that only a sample of the population has been selected, and a different sample would probably produce a different estimate. Non-sampling errors are covered in section 7 of this volume. They measure error that is not due to random sampling and would remain even if every case in the population was sampled.

Sampling errors relate to the fact that the sample chosen is only one of a very large number of samples which might have been chosen. It follows from this that an estimate of, say, the number of people in employment, is only one of a large number of such estimates which might have been made. The probable spread of these different estimates can be thought of as the precision of our estimate of the number of people in employment. Greater precision is associated with a relatively narrow spread. In general, a larger sample size equates to a smaller spread of probable estimates and therefore greater precision.

8.1 MEASURING PRECISION USING STANDARD ERRORS

A measure of the spread of different probable estimates is provided by their *standard error*. This is the *standard deviation* (the average amount of variation about the average) of the estimates which would have arisen from the different samples which might have been taken. The smaller the standard error, the more precise is the estimate.

The size of standard errors is determined by a number of factors, including the sample size and the variability of the population from which the sample is drawn. The third important factor in determining the size of standard errors is the sample design. Standard errors calculated from simple random samples will differ from those calculated from more complicated sample designs, such as clustered or stratified samples.

A useful benchmark to assess the relative magnitude of a standard error is to calculate the ratio of the standard error derived from a particular (complex) sample design with the standard error that would have arisen from a simple random sample of the same size. This ratio (of the standard errors) is the *design factor*. It indicates the relative gain (or loss) in the estimate of standard error which results from the use of a particular complex sample design compared to a corresponding simple random sample. A design factor (or DEFT) of, say, 1.20 indicates that the standard error of the estimate in question is 20% greater than would have been the case for a simple random sample of the same size. The design factor (DEFT) should not be confused with the design effect (DEFF); the design effect is the design factor squared and is calculated by the ratio of variances instead of standard errors.

In the case of the LFS sample design, there is a clustering effect. This reflects the fact that addresses are sampled, but that results are shown for individuals. For example, ethnicity is particularly clustered, since it is likely that all members of a household living at a particular address will share the same ethnicity. This results in,

for example, the The design factor for part-time employees on the other hand is 1.12, reflecting the fact that part-time employee status is not clustered within a household.

By itself, clustering would tend to increase the design effect of LFS estimates. However, the LFS uses a systematic sample of addresses ordered by postcode sector (see section 3). This means that the sample will be representative geographically, which will reduce standard errors. This effect is referred to as *implicit* stratification (as opposed to explicit stratification, where a population is split into sub-groups and a sample is taken from each sub-group). The weighting process used on the LFS will also reduce standard errors because it ensures that the weighted dataset is representative of the population (see section 10).

8.2 CALCULATING CONFIDENCE INTERVALS

While the standard error and the design factor of estimates are important items of information in their own right - because they indicate the precision of the estimate and the relative efficiency of the sample design in deriving the estimate - they also form the basis for calculating confidence intervals associated with particular estimates. A 95% confidence interval for a population estimate is ±1.96 standard errors around the estimate calculated from the sample.

The tables in Annex C to this volume of the LFS User Guide, list a number of estimates from the January 2023 to December 2023 LFS, the standard errors and other related st

atistics. The standard errors presented in the Labour Market Statistics statistical bulletin, LFS Quarterly Supplement and Annex B are shown for the UK and are calculated using UK design factors. For example, in October-December 2023 the proportion of people in the UK aged over 16 who were estimated to be economically active was 0.64. The number of people aged 16 and over in the UK sample was 64,481 and since the design factor was 0.95, the standard error was 0.0018, calculated as:
SE (given LFS design) = Design Factor * SE (assuming simple random sample)

$$= \text{DEFT} * \sqrt{\frac{p(1-p)}{n}}$$

$$= 0.95 * \sqrt{\frac{0.64*0.36}{64,481}} \qquad = 0.0018$$

Hence a 95% confidence interval would be:
        0.64 ± (1.96 * 0.0018) = 0.64 ± 0.0035 = (0.6365, 0.6435)

What this means in practice is that in 19 samples out of 20 we would expect the true economic activity rate to lie within the 95% confidence intervals constructed. Only in 1 in 20 samples would we expect the true value of the economic activity rate to be outside the confidence interval around the LFS estimate.

The standard error of the total of the estimate is approximately the standard error of the proportion (or rate) multiplied by the population aged 16 and over:

0.0018 * 53,415,389 = 96,147

The standard error of the total of the estimate (i.e. 96,147) multiplied by 1.96 gives the margin of error:

96,147*1.96 = 188,448

The number of economically active people aged 16 and over is: 34,250,854.

Therefore, the 95% confidence interval for the number of economically active people aged 16 and over would be:

34,250,854 ± 188,448

Note that this method for constructing standard errors for totals can only be used when the base (the total the proportion is based on) is not itself a survey estimate. In practical terms, this means that this method should only be used when the base is a weighting category or a combination of weighting categories – see section 10.2, and is inappropriate when using the income weight – see section 10.5.

We use 95% confidence intervals primarily because they are widely used within ONS and elsewhere; they are something of an industry standard. However, other confidence intervals are equally valid, and might help users appreciate the fact that LFS estimates are always subject to error.

8.3 CALCULATING LFS STANDARD ERRORS & CONFIDENCE INTERVALS

The standard errors of the UK LFS estimates shown in Annex B are produced using a linearised jackknife approach in order to account for the impact of the calibration weighting on the variance of estimates - calibrated variances will tend to be smaller than non-calibrated variances because calibration ensures that estimates are representative across calibration categories[19]. Paired addresses (sorted by wave, quarter, quota, week and address number) are used as stratifiers and the address is the primary sampling unit (PSU). This is also the method used for standard errors produced for the first release. It does not take into account the impact of seasonal adjustment on standard errors.

The design factors given in the appendix can be used with the formula given in section 8.2 to generate standard errors for key labour market estimates. If other estimates are required, users may wish to calculate their own standard errors. In this case, complex survey commands in statistical software packages can be used – this will account for the survey design but will not account for the reduction in standard errors resulting from calibration weighting, and so this method is likely to produce conservative standard errors. The address identifier should be used as a cluster variable, as LFS person responses are clustered within addresses. Because the LFS

---

[19] Holmes, D. J. and Skinner, C. J. (2000), "Variance estimation for Labour Force Survey estimates of level and change", *No 21, Government Statistical Service Methodology Series,* available at http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/method-quality/specific/gss-methodology-series/index.html

uses systematic sampling, a good geographic coverage is ensured – this is called implicit stratification (see section 8.1). This can be accounted for by using a low level of geography as a stratum variable.

In the case of totals based on small sample sizes, a useful approximation for the 95% confidence interval of an estimate of 'M' is:

$$M \pm \left[1000 * \sqrt{\frac{M * 1.92}{1000}} \right]$$

The derivation of this formula is given in appendix B. It implies that, for an LFS estimate of 20,000, the confidence interval will be approximately –

$$20,000 \pm 1000 * \sqrt{\frac{20,000 * 1.92}{1000}} = 20,000 \pm 6197.$$

This method may not work for estimates of total employed and total inactive as these tend to have very low design effects due to the effect of post-stratification.

ONS Methodology have produced a guide to calculating standard errors, which includes information on how to use statistical software to calculate standard errors that account for the survey sample design survey in addition to the sample size and the variability in the population:
https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/onsmethodologyworkingpaperseriesno9guidetocalculatingstandarderrorsforonssocialsurveys

8.4 STANDARD ERRORS AND CONFIDENCE INTERVALS OF ESTIMATES OF CHANGES

In the same way that standard errors relating to quarterly estimates of means or proportions can be calculated, so standard errors can be calculated which relate to changes.

Standard errors (and hence confidence intervals) of estimates of changes are calculated as follows; where p_{t1} is the relevant proportion at time 1, p_{t2} is the relevant proportion at time 2, k is the sample overlap (approximately 0.8 for successive quarters) and r is the correlation coefficient -

$$\text{var}(p_{t1} - p_{t2}) = \text{var}(p_{t1}) + \text{var}(p_{t2}) - \frac{2kr[\text{var}(p_{t1}) + \text{var}(p_{t2})]}{2} \quad (1)$$
$$= [\text{var}(p_{t1}) + \text{var}(p_{t2})](1 - rk)$$

$$se(p_{t1} - p_{t2}) = \sqrt{\text{var}(p_{t1} - p_{t2})} \quad (2)$$

This is the method used for standard errors for the main labour market release, with correlation coefficients calculated as detailed in Holmes and Skinner (2000). With the additional assumption that $\text{var}(p_{t1}) = \text{var}(p_{t2})$ it can be simplified further -

Since $\text{var}(p) = deft^2 * \dfrac{p(1-p)}{n}$, if $\text{var}(p_{t1}) = \text{var}(p_{t2})$ then –

$$\text{var}(p_{t1} - p_{t2}) = deft^2 * 2(1-rk) * \left[p(1-p)/n\right] \qquad (3)$$

$$se(p_{t1} - p_{t2}) = deft * \sqrt{2(1-rk)\left[p(1-p)/n\right]} \qquad (4)$$

A few key labour market standard errors for two-quarter changes in totals are given in the table below, calculated using formulae (1) and (2) :

Table 8.1: Standard error for two-quarter changes for economic activity status, Q1 2010 to Q2 2010

Q1 2010 to Winter Q2 2010 (UK) (not seasonally adjusted)	Quarterly correlation	Quarterly change	Standard error of change in level
Economically active	0.8	120,526	61,886
Employed	0.8	188,247	65,663
ILO unemployed	0.54	-67,721	42,879
Economically inactive	0.8	-21,544	61,886

As the SE for change in total employment is given as 65,663; then a 95% confidence interval for the change of +188,247 can be calculated as:
$$188,247 \pm (1.96 * 65,663) = 188,247 \pm 128,700$$

8.5 RELIABILITY THRESHOLDS

It is the nature of sampling variability that, when estimating population sizes, the smaller the group whose size is being estimated, or from which an estimate is being derived, the less precise that estimate is. Put another way, the size of the standard error increases with the level of the estimate, so that the larger the estimate the larger the standard error. But the larger the sample estimate, the smaller will be the standard error in percentage terms (relative standard error being the standard error as a percentage of the estimate). Thus, larger sample estimates will be relatively more reliable than smaller estimates–an estimate of 500,000, while having a standard error of 13,800 will have a relative standard error of 3%, compared with an estimate of 25,000 which has a standard error of 3,100 and a relative standard error of 12%. This is because larger estimates will be based on more cases.

Before 2005, quarterly LFS estimates of under 10,000 were not published as they are likely to be unreliable. Although this publication policy changed in 2005 (as a result of the Freedom of Information Act), the unreliability of these LFS estimates did not. It is suggested that the 10,000 threshold is used as a guide to identify which cells will be subject to high sampling variability. Users are advised that estimates

below 10,000 are subject to a high degree of sampling variability and should therefore be treated with caution. The 10,000 threshold equates to a sample size of about 25 and a relative standard error of about 20%. The graph below shows how, for different values of *p* (i.e. for different proportions of the population), the standard error rises at a much steeper rate when the sample size is less than 30 - very small estimates, those based on fewer than 25 cases, are subject to such high standard errors (relative to the size of the estimate) as to detract seriously from their value, which is why the reliability threshold is recommended to be around that level. That said, a relative standard error of 20% is a somewhat arbitrary cut-off point, the Australian Bureau of Statistics uses a publication threshold of a relative standard error of 25%. A sample size of fewer than three is potentially disclosive so this information is suppressed even under the Freedom of Information Act.

Variation in standard errors according to changes in p and sample size (n)
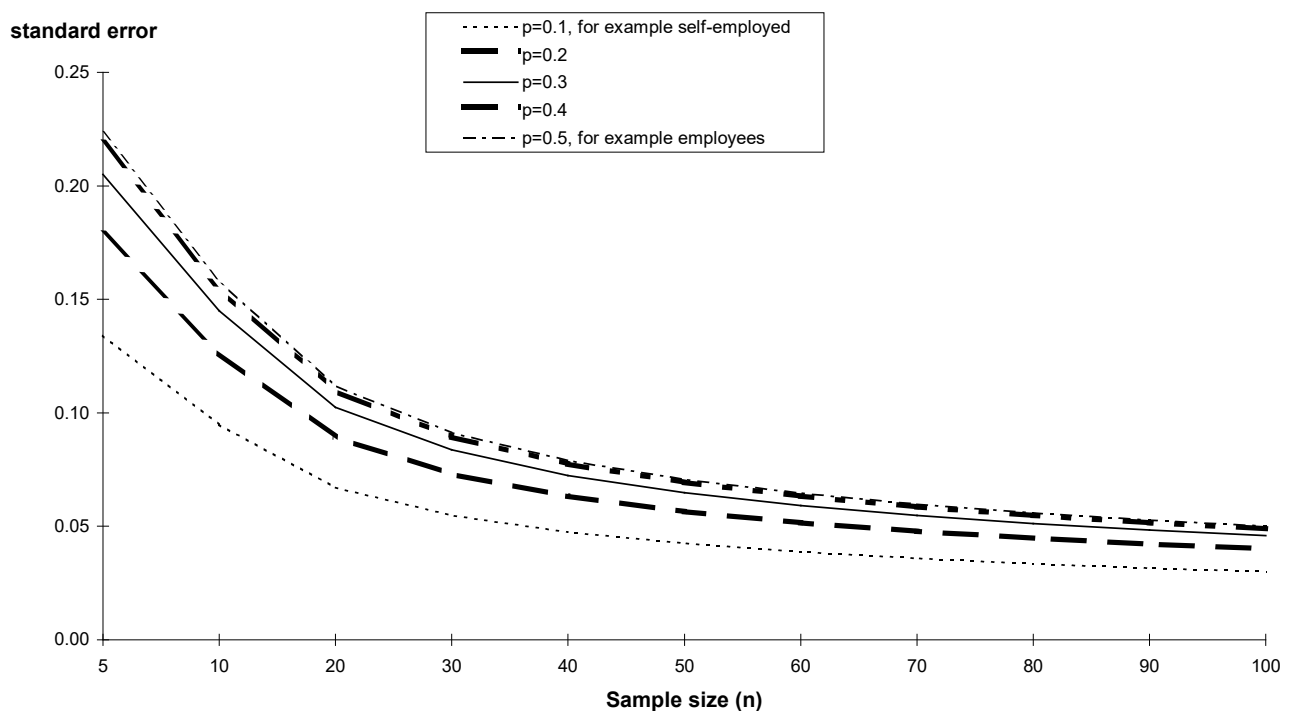
Estimates of 10,000 have associated 95% confidence intervals of 10,000 ±4,000 -so particular care must also be taken in using estimates of change from period to period for these estimates. For example, the estimated size of a particular group may be unchanged at 10,000 in two successive quarters but, because of the unreliability of these estimates, we cannot safely conclude that the true size of the group has not changed between those quarters. A variable with a quarterly correlation of 0.9, would have a 95% confidence interval for its quarterly change of 0 ± 3000. So, even if the estimate does not change, the true size of the group could have changed by up to 3000 in either direction.

8.6 RELIABILITY THRESHOLDS FOR EARNINGS DATA

For estimates of the number of people in a small group, which is a count, for example employed people in a small ethnic group, we can use an approximation of the variance to derive the minimum number of cases that is required in a group to achieve a relative standard error of less than 20%. However, Earnings cannot be

regarded as a count, it is a continuous variable, and hence the method for counts does not apply. There is no approximation method that can be used to derive a reliability threshold of variables that are not counts. Instead, we propose a threshold based on values of relative standard errors of small groups that were computed using recent APS earnings data.

Relative standard errors were obtained for estimates of mean gross earnings for groups defined by UALAD and age (grouped) and by UALAD and ethnicity (grouped). In both sets of groups, all groups with 25 or more cases had a relative standard error less than 20%. On the other hand, in groups with fewer than 25 cases, a proportion of the groups had a relative standard error higher than 20%. Estimates of counts also have a reliability threshold of 25 cases per group.

The threshold depends on the variation of earnings, the sample design and weighting method, and hence may need to be revised in the future. We, therefore, recommend to use a reliability threshold of 25 cases for estimates of earnings and monitor its value regularly, every two years, for example.

8.7 RELIABILITY THRESHOLDS FOR ANNUAL LFS DATA

For Annual LFS data prior to 2000-01 when the survey was enhanced, estimates of fewer than 6,000 are likely to be unreliable. However, since 2000-01, the nature of LFS enhancement has meant that some areas have seen a very large increase in sample size, and others a very small increase or none at all. This means that a single threshold for all areas is no longer appropriate.

Following the enhancement in 2000-01, each area in England was allocated to one of three threshold bands – 2,000, 4,000 or 6,000. Annex D of Labour Force Survey User Guide Volume 6 Local Area Data, contains details of how this allocation has been made.
For Wales, from 2001-02, each Unitary Authority was allocated to one of four threshold bands - 1,000, 2,000, 3,000, or 4,000.
Similarly for Scotland, from 2003-04 each Unitary Authority was given its own threshold ranging from 1,000 to 5,000..
It has long been known that the design effects for ethnic group and for totals segregated by ethnic group are substantially greater than those for most other groups. Following an analysis of the variation in design effects between different ethnic groups and different local areas in England, it is recommended that for most ethnic estimates a single multiplier of 2.5 is applied to the general thresholds. A larger multiplier of 3.5 is recommended for totals of individual minority ethnic groups (e.g. the multiplier of 3.5 would apply to the total Indian adults in Birmingham, whereas a multiplier of 2.5 would apply to the total employed Indian adults in Birmingham). A separate analysis for the Welsh Local Labour Force Survey recommended a multiplier of 4.0 in Cardiff and 2.5 for the rest of Wales.

SECTION 9 – NON RESPONSE

This section looks firstly at the results from the Census Non-Response Link Study (CNRLS) conducted on the LFS following the 2011 Census and then at question specific non-response.

9.1 THE 2011 CENSUS NON-RESPONSE LINK STUDY

Since ONS carry out both the Census and the LFS, a study could be carried out, which involves linking census records with the LFS wave 1 cases (in England and Wales)[20] that were sampled between March and July 2011 (in Q1 and Q2). This allows a comparison to be carried out on the characteristics of survey respondents and non-respondents, which enables non-response bias to be estimated. Methods of non-response adjustments could then be evaluated to see whether a non-response weight should be introduced on the LFS.

Similar work was carried out following the 1971, 1981, 1991 and 2001 census[21] , for these periods, it was found that the addition of non-response adjustment to the existing weighting methods only had a negligible effect on estimates and therefore wasn't implemented.

The results of the census non-response link study can be found in the paper "Non-response weights for the UK Labour Force Survey? Results from the Census Non-response Link Study".
http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/articles-and-reports/index.html

For the 2011 CNRLS, 12,790 households from wave 1 LFS were successful matched to the census records, of these: 62.2% co-operated, 11.6% were non-contacts and 26.2% were refusals. Four logistic regression models were used in the study (univariate model (ethnicity-based model), multivariate, tree and quality-check). Details of these models and the results can be found in the report.

The conclusion of the study was not to introduce a non-response weight on the LFS at this time, as:
   a) non-response adjustments have a negligible impact on estimates for the key publication groups explored, except for Ethnicity. However, while the impact on levels within minority ethnic groups is notable, the impact on rates is negligible.(The number of respondents in ethnic groups other than "White" are rather small, which means that the calculated factors from the CNRLS data may not be very reliable. In the "white ethnic group" non-response adjustment had no impact).
   b) response rates have continued to decline which indicates that the response process isn't stable as shown in section 5 of this user guide and in the PQMs. This can also be seen when comparing the response rates in 2001 and 2011 as in 2011 response rates of unemployed (62.1%) were higher than the employees (61.3%) whereas in 2001 response rates were higher for employees (81.1% compared to 74.6%) as shown in the CNRLS report.

---

[20] Linked census-survey data from Scotland were unavailable at the time of the analysis.
[21] Information on the 2001 census linked study can be found in the 2011 version of this user guide.

9.2 QUESTION SPECIFIC NON-RESPONSE

Sometimes data for a particular question is missing from the results because respondents who are routed to the question do not respond. The figures in table 9.1 represent the number of people who did not answer particular questions as a percentage of those who were routed to the question.

Table 9.1 Question specific non-response rates

Variable	OD22	JM23	AJ23	JS23	OD23	JM24
EVERWK (Whether ever had a paid job)	0.05	0.07	0.05	0.03	0.03	0.03
EVEROT (Paid or unpaid overtime)	0.19	0.18	0.14	0.15	0.19	0.14
FTPTWK (Whether working FT/PT)	0.12	0.11	0.08	0.09	0.09	0.06
ILLWK (Whether off sick or injured in ref week)	0.00	0.00	0.00	0.00	0.00	0.00
SECJOB (Second Job)	0.13	0.07	0.05	0.04	0.08	0.04

Users should be aware of this when undertaking analysis.

To conclude, various strategies are available to weight survey data to take account of non response. A census based weighting scheme is one approach. ONS currently uses another approach, a population weighting procedure, which involves weighting data to sub-regional population estimates and then adjusting for the estimated age and sex composition by region. When evaluating whether additional information might improve the performance of a population weighting methodology, it is important not to lose sight of issues such as the availability of up-to-date population data at the sub-regional level and the relative transparency of the method.

SECTION 10 - WEIGHTING THE LFS SAMPLE USING POPULATION ESTIMATES

The LFS collects information on a sample of the population. To enable us to make inferences from this sample to the entire eligible population we must weight the sample data. This entails assigning each responding or imputed case a weight, which can be thought of as the number of people in the population which that case represents. These weights are calculated such that they sum to a set of known population totals, and the weights of an entire dataset will sum to the eligible population of the UK.

Population weighting serves several purposes. It ensures that estimates reflect the sample design so that cases with a lower probability of selection will receive a higher weight to compensate. It also compensates for differential non-response among different sub-groups in the population, and as such should help guard against potential non-response bias. The use of weights also allows totals, as well as means and proportions, to be estimated easily; and weights may reduce standard errors when the calibration model is included in the variance estimator.

It is therefore important to use the weights when doing most types of analysis on LFS datasets. Failing to do so may introduce bias because the sample design will not be taken into account – for example, over-75s will be under-represented, as they are under-sampled. Not using the weights will also result in estimates that are subject to more non-response bias and will make it difficult to estimate totals.

The person-weight variable on LFS datasets is usually named 'pwtXX', where 'XX' refers to the year in which the population totals were projected (see section 10.3). This is the main weight used for inference to person-level population characteristics such as economic activity rates and totals. A household weight and an income weight can also be found on some datasets and may be more appropriate for some types of analysis; these are described briefly in sections 10.4 and 10.5.

10.1 CALIBRATION WEIGHTING THEORY

The LFS uses calibration weighting to assign a calibration weight w_k to each responding individual *k*. These calibration weights are set to sum to a set of *calibration totals* within *calibration groups* – for example, the weights of all 18-year old males in an LFS dataset (a *calibration group)* will sum to the population total of eligible 18-year old males in the UK (a *calibration total*) at the time the survey was taken.

Calibration weighting typically involves calculating a design weight, making adjustments for non-response, and finally calibration to population totals. The design weight d_k for each individual *k* is calculated as the inverse of the probability of selection p_k, so that individuals with a lower probability of selection receive a higher design weight.

$d_k = 1 / p_k$

It is possible to modify the design weight by non-response factors in order to account for some sub-groups being less likely to respond than others. Non-response factors are not currently used on the LFS but they may be introduced in the future. See section 9 for more details about non-response.

It is desirable for the calibration weight to be as close as possible to the design weight, in order to properly reflect probabilities of selection. The calibration weights $w_k = d_k g_k$, $k = 1,....,n$ are calculated to minimise the sum of the distances between the d_k and the w_k, subject to the calibration constraints (ie summing to the known totals).

In summary - the calibration weight is calculated to sum to calibration totals within calibration groups while minimising the adjustment to the design weight.

10.2 CALIBRATION WEIGHTING ON THE LFS

The LFS assigns a calibration weight to all responding or imputed individuals but does not assign a weight to individuals whose economic activity is unknown (so non-responders do not get a weight). Standard LFS practice in the case of individuals dropping out between waves is to roll their data forward by one quarter – this is a form of imputation, and these individuals receive a weight.

LFS design weights are typically constant, as in most cases the LFS sample design ensures an equal probability of selection – see section 3. There are some exceptions to this ;

- A different sampling fraction, and therefore a different design weight, is used in Northern Ireland.

- From Q3 2010 onwards, where multiple households are resident at a sampled address, only one household is interviewed (see section 3.3.1 for more details). The probability of selection for households in multiple-household addresses is therefore lower and their design weight is higher.

- From Q1 2020 to onward, to account for the low representation of the over-65s a scaling factor is used to adjust the design weights. Previously from Q3 2010 to until Q4 2019, only over-75s design weights were adjusted as they are only interviewed in Wave 1 (see section 3.4.1 for more details).

- From Q1 2020 to onward, to account for potential bias from the change in data collection a non-response adjustment is applied to design weights.

These design weights are then calibrated to sum to population totals using the theory outlined in 10.1. Five sets of calibration groups (called partitions) are used, so that each individual is in five separate calibration groups. Recall that, within each

calibration group, weights will sum to population totals. The calibration groups used are:

Partition 1:   Individual Local Authority Districts

There are 433 local authorities in total, meaning there are 433 calibration groups in this partition.

Note: Due to insufficient sample size (required for smooth running of the weighting process), some local authorities may need to merge with others which can reduce the number of calibration groups in this partition.

Partition 2:   GB/NI by sex for the ages 0-15, 16-19, 20-24 and 25+

This partition is a cross-classification, so that age-bands within sexes and GB/NI are used to form calibration categories. As there are 4 age groups, two sexes and two countries in this partition, there are 4*2*2 = 16 calibration groups.

Partition 3:   Male/Female by Government Office Region (GOR9D) and Age-Groups -

GOR9D: North East   North West& Merseyside

Yorkshire and Humberside

West Midlands   East Midlands

East of England   London

Sout East   South West

Scotland   Wales

Northern Ireland

Age groups: 0-15   16-24   25-34   35-44   45-54   55-64   65-74   75+

Like partition 2, this partition is a cross-classification, so that age-bands within government office regions and sexes are used to form calibration groups. As there are 12 regions, 8 age-bands and two sexes in this partition, there are 12*8*2 = 192 calibration categories.

Partition 4:   GB/NI by sex for the age groups 0-4, 5-9, 10-15, 16-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75+. This partition is similar to partition 2 but with more detailed age-group information. As there are 16 age-groups, two sexes and two countries in this partition, there are 16*2*2 = 64 calibration groups.

Partition 5: Tenure groups. This partition is used to ensure that the distribution of the weighted achieved sample is uniform over the census 2021 tenure distribution. As there are 3 main tenure groups (Owned outright, mortgage or loan and Rent) used in weighting process, meaning there are 3 calibration groups in this partition.

There are therefore 3 + 64 + 192 + 16 + 433 = 708 total calibration groups, and each responding/imputed individual will be in five of the partitions. For example, a 24-year-old man in Cardiff would be in the Cardiff calibration category in partition one, the 24-year-old male in GB category in partition two, the Welsh 16-24-year-old male category in partition three, 20-24-year-old male in GB category in partition four and 20-24-year-old male in respective category of tenure in partition five. The weights of all calibration categories will then sum to corresponding population totals – for example, the weights of all responders in the Welsh 16-24-year-old male calibration category will sum to the Welsh 20-24-year-old male population total.

The LFS is calibrated using Statistics Canada's Generalised Estimation System (GES) software, a set of programs designed for calibration weighting. Prior to 2007/8, a raking ratio method was used instead. Raking ratio sets the weights to sum to the calibration totals for each partition in turn – the first stage corrects for partition one, the second for partition two, and the third for partition three. This is then repeated iteratively until the weights sum (sufficiently closely) to all population totals in all three partitions. This method was replaced by GES in 2007/2008 and GES is viewed as more statistically robust and efficient, calibrates to all partitions simultaneously, and allows good variance estimates. GES was used to re-weight back to 1991/1992, so all post-1992 LFS datasets now available will have a weight calculated using GES.

The 2011 census resulted in revisions to ONS population estimates, and in 2014/2015 a re-weighting program was carried out to re-weight historical datasets to updated population totals back to 2001. In 2016 another re-weighting exercise took place, going back to 2012. In 2021 an interim re-weighting exercise took place, going back to 2020. Another re-weighting exercise related to population changes took place in 2023 going back to 2022. To account for the changes in the population estimates related to 2021 census including Scotland and Northern Ireland, ONS is planning to run a full-scale reweighting exercise back to 2011 in 2025.

10.3 POPULATION TOTALS AND RE-WEIGHTING

The LFS weighting methodology requires estimates of the number of people in LFS-defined households[22] for each local authority, with a five-year age-breakdown by sex. Population weighting totals are derived from published population estimates and population projections. All population projections (and estimates) are based, directly or indirectly, on the decennial Census of Population, and use additional information from the NHS Central Register for internal migration, the International Passenger Survey for international migration flows, and registration data for births and deaths. See the Quality Monitoring Information (previously, Summary Quality Reports) for

---

[22] Household residents and NHS staff and students

Mid-Year Population Estimates and National Population Projections for more details[23]. Projections use a variety of assumptions about the rates at which the components of population change will evolve.

A number of adjustments are made to the published population estimates by local authority and national population projections data to provide estimates of the number of people in LFS-defined households:

(i)   Population projections for local authorities are produced by rolling forward a five-year average growth rate (between estimates) for each LAD, then constraining to the published national population projections.

(ii)   Estimates of communal establishment population (that is, those excluded from the LFS-defined population) have been made by assuming that the percentage of people in communal establishments was the same (by quinary age band, sex, and region) as it was in the results of the 2011 Census. Hence, for example, as the number of old people changes, the number of old people in institutions changes in line. Then the LFS-defined population is calculated by subtracting the estimate of communal establishments from the total population figures. See section 3.1.2 on Communal Establishments for more details.

(iii)   Monthly estimates are produced from the LFS-defined annual population totals - the mid-year estimate/projection less the communal establishment population - by simple linear interpolation.

Re-weighting exercises give the opportunity to use the most up to date population estimates as a part of LFS weighting. In the interests of timeliness, population projections are used to generate the initial population totals for weighting. Population estimates are published one year following the reference date and it is desirable to re-weight to the resulting LFS-defined population totals given by these published estimates. Re-weighting consists of re-calibrating historical data, which had initially been weighted to population projections, to newly available population estimates, and then re-projecting into the future. The year in which the population projections have been updated using the most recent population estimates is reflected in the weight variable – for example, 'pwt18' is based on projecting from 2018 population estimates. A 2017 dataset with 'pwt18' is therefore weighted to 2018 population estimates calculated in 2017.

In the light of the pandemic and the introduction of a mode change (from face-to-face to telephone in wave 1) due to the data collection constraints an adjustment to the weight calibration was introduced using the 'tenure' variable.[24]
Furthermore, in the absence of regularly published population estimates and projections, a population growth rate based on the HMRC PAYE Real-Time

---

[23]https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/previousReleases
[24]
https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/coronavirusanditsimpactonthelabourforcesurvey/2020-10-13

Information (RTI) data was used for an interim reweighting exercise in 2021 and 2022 covering all different types of LFS and APS outputs.[25]

Following the publication of National Population Projections (NPP) based on the 2021 Census for England and Wales in 2023, a further interim reweighting exercise was conducted, which saw the LFS person level weights being re-released on this new basis in February 2024 including the periods July-September 2022 to July-September 2023. From October-December 2023 onwards the LFS person as well as the income weight are on the new NPP basis.[26]

ONS is considering a number of options for another reweighting exercise and will inform users about further up-dates going forward.

10.4 THE LFS EARNINGS WEIGHT

Earnings data are collected at Wave 1 and 5 interviews only (before Spring 1997 at Wave 5 only). A separate weight is needed for analysis because data are only collected at two of the five waves - using the normal person weight would be inappropriate, as the weights would sum to considerably less than the relevant population total. During 1998, income weights for Northern Ireland data were added to existing datasets so that it is possible to analyse earnings data at UK level from Winter 1994/5 onwards. However, the weighting exercise is restricted to employees' earnings, and people with a very high income (over £3500 gross per week) are treated as outliers and do not receive an income weight. Individuals where data is brought forward from a previous quarter will also not receive a weight (because the questions are only asked at waves 1 and 5).

The aim of the income weight is to allow inference from the income data to the entire target population. This target population is different to the target population which the normal weight refers to because only employees are eligible for earnings questions. The aim is therefore to weight the earnings data to allow inference to the UK employee population. The best source of information of the size of sub-groups in this employee population is considered to be the full LFS dataset, containing all five waves of data. The weighting procedure therefore attempts as far as possible to replicate the results from the employees in waves 1-5 of the LFS in the weighted earnings data, and as such can be thought of as a form of two-phase weighting. It is worth noting that as the total size of the target (employee) population has to be estimated from LFS data, standard errors will be larger than those from the main LFS. This is because the population totals used in the weighting are themselves estimated from the survey.

A small number of variables that are likely to be important determinants of income are used to form calibration categories. GES is used to calibrate the weights of

[25] https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyweightingmethodology

[26] https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/impactofreweightingonlabourforcesurveykeyindicators/2024

income respondents within these categories to sum to the totals estimated by the weighted full LFS sample of employees. The calibration groups used are -

Partition 1: Five-year age bands by sex
There are 10 age bands and each is broken down by sex, so there are 10*2=20 calibration groups in this partition.

Partition 2: Full-time/Part-time by Standard Occupational Classification Major Group
There are 10 Standard Occupational Classification Major Groups and each is broken down by full time/part time, so there are 20 calibration groups in this partition.

Partition 3: Standard Industrial Classification Industry Sector
There are nine Standard Industrial Classification Industry Sectors, so there are nine calibration groups in this partition.

Partition 4: Government Office Region (summary) by sex
A nine-group aggregation of Government Office Region is used, showed as GOVTOF2 on LFS datasets. There are nine such summarised Government Office Regions and each is broken down by sex, so there are 18 calibration groups in this partition.

Industry and Occupational classifications are periodically updated, and the income weighting is altered to reflect this – since January 2011 the income weighting has used SOC 2010, and since January 2009 it has used SIC 2007.

10.5 OTHER WEIGHTS

10.5.1 Household

The primary use of the LFS is producing person-level statistics (such as employment, unemployment and economic inactivity levels and rates) broken down by personal characteristics (such as age, sex and region). However, as the survey collects information about all eligible individuals at responding households, it is also possible to produce person-level characteristics broken down by the characteristics of households in which people live, and to estimate the total number of households of a particular type (for example, the number of workless households). These types of estimates can be thought of as household-level estimates.

Using the weight described in section 10.2 (the 'person-weight') to do household-level analysis can lead to biased estimates. For this reason, separate LFS datasets designed for household-level analyses are produced for every quarter of each year. A description of these datasets and problems of bias in household-level analysis can be found in volume 8 of the LFS User Guide.

The household dataset uses a different weighting methodology to the person-level dataset. There are three main differences –

1) All individuals in households with at least one respondent receive a weight. This means that non-responders in partially responding households receive a household weight but not a person weight.
2) The household weight is calculated using integrative calibration, which means that all members of a household receive the same weight, which can also be regarded as the weight for that household.
3) A slightly less detailed set of calibration categories is used.

More details of the household weighting methodology can be found in volume 8 of the LFS user guide.

10.5.2 Annual Population Survey

The Annual Population Survey (APS) is a composite of LFS and APS boost data published over periods covering a year. For details about the APS datasets and their use see volume 6 of the LFS user guide and section 3.6 of this volume.

The APS weighting uses the same basic methodology as the LFS weighting - both the APS and LFS weights are calibrated to population totals using GES. The main difference is that, as more cases are available, it is possible to use a more detailed set of calibration groups when weighting the APS. Additionally, because of the APS sample design, the design weight used as an input to the weighting will vary between local authorities. Like the LFS, the APS design weight is not adjusted for non response before it is calibrated (see section 10.1).

A household-level APS dataset is also available for calendar years. This dataset uses a similar weighting methodology to the household-level LFS dataset, but with a more detailed set of calibration groups. For more details, see volume 8 of the LFS user guide.

An APS earnings weight is also available for the calendar years (January-December periods) from 2012 and for all periods from 2018. Again this uses similar weighting methodology to the LFS earnings weight, but with a more detailed set of calibration groups (6 rather than 4). The calibration groups are:

Partition 1: Grossing area by Five-year age bands.
There are 9 age bands and 46 grossing areas giving 414 partitions
Partition 2: Local Authority area
where areas with low numbers of responses are combined with other local authority giving approximately 403 partitions
Partition 3: Government office region by five-year age bands by sex.
The government office regions are in 13 groups resulting in 234 partitions.
Partition 4: Full-time/Part-time by Standard Occupational Classification Major Group
There are 9 Standard Occupational Classification Major Groups and each is broken down by full time/part time, so there are 18 calibration groups in this partition.
Partition 5: Standard Industrial Classification Industry Sector
There are nine Standard Industrial Classification Industry Sectors, so there are nine calibration groups in this partition.

Partition 6: Sex by five-year age bands
for over 25 year olds and yearly age bands for 16 to 24 giving 21 age
bands and 42 partitions.

More information on the APS earnings weight including the main uses can be found in the volume 6 user guide.

10.5.3 Longitudinal Analysis

The LFS was designed to produce cross-sectional data, however, it is recognised that linking together data on individuals across quarters can produce a rich source of longitudinal data. To this end, longitudinal datasets are made available with individuals linked across two or five consecutive quarters. Two-quarter longitudinal datasets contain individuals who responded in two consecutive quarters and include their responses at each quarter, and five-quarter longitudinal datasets link individuals who responded in five consecutive quarters (i.e. – who responded in waves 1-5).The two year APS longitudinal dataset link individuals who responded in two consecutive January-December APS datasets .

Longitudinal datasets are weighted to allow inference to the eligible population totals in a similar fashion to the person and income weights. A known issue with longitudinal data analysis is that it can be biased by differential attrition – some groups of people are more likely to drop out of the survey between quarters than others. The longitudinal weighting is designed to help guard against this kind of bias.

Longitudinal weighting entails calibration of a linked datasets to population totals with additional adjustments to help combat attrition. A summary of the weighting process can be found in the longitudinal user guide (volume 11), with more detail available in a published report in the *Government Statistical Service Methodology Series 17*[27]. The report also provides more detail about the methodological development of the longitudinal datasets.

---

[27] Clarke, P S & Tate, P F (1999) 'Methodological issues in the production and analysis of longitudinal data from the Labour Force Survey', *Government Statistical Service Methodology Series* 17*,* available at
http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide -method/method-quality/specific/gss-methodology-series/index.html

SECTION 11 - REPORT ON PROXY RESPONSE STUDY BASED ON LFS QUESTIONS

In order to maximise response in the short fieldwork period available and to contain the cost of recall interviews, interviewers are allowed to accept information by proxy for those household members not available when the interview takes place (see section 5.5 for more information).

Martin and Butcher (1982) 'The Quality of Proxy Information - Some Results from a Large-scale Study', *The Statistician,* Vol 31, No.4 showed that the use of proxy responses has a greater effect on some variables compared to others. In 1995, Social Survey undertook a study on key LFS variables including economic activity, hours worked, second job, income, training, and education. The report "A study of proxy response in the Labour Force Survey" was written by Fiona Dawe and Ian Knight and was published in the Survey Methodology Bulletin (No.40), January 1997. *http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/method-quality/survey-methodology-bulletin/smb-40/index.html*

The report concluded that:
- The agreement levels between information given by proxy informants and the same information given by the subjects themselves were for many key variables above 80% and several were above 90%. Taking account of the proportion of LFS responses given by proxy, this implies overall gross error rates of around 1-5% for these variables.
- Those variables requiring less straight-forward information (such as training in the last week and highest qualification obtained) and those requiring very detailed numerical information (such as hours worked and income) showed a less matches between proxy and subject responses which means higher gross error rates. For example, for gross weekly income, 66% of proxies were able to give an answer within 10% of the subject's answer and if such data were taken from spouse proxies only, the level of matching would have been much higher. A bigger problem for such variables is the high level of "don't knows."
- The reliability of proxy data, for areas where the proxy informant is required to provide precise numerical answers (e.g. hours worked and income), increases with a decrease in the required level of detail. If the data is used in a banded form, or to calculate averages the match of proxy and subject data is greatly increased.

The study also investigated whether it was possible to identify a key relationship of proxy to subject that would guarantee a lower rate of proxy error. Whilst spouse proxies were better for some variables like occupation they were worse for others like qualifications. In general, no single type of household member is able to supply reliable proxy information for all questions, it would seem that the best placed household member to provide proxy information is the person most affected by the subject's actions, though the income variables would be more reliable if restricted to spouse proxies.

SECTION 12 - IMPUTATION IN THE LFS

In the earlier section regarding dependent interviews it was noted that for many quarters, responses may be rolled forward (for one quarter only) if a respondent is unavailable. This is referred to as 'imputation'.

The following examines some of the implications of this. The first part deals with the situation that arises in the case of non-core questions (which are not asked in every quarter). If a respondent is unavailable in the latest quarter, then the variable will be coded as DNA (Does Not Apply – there will be no data to 'roll forwards'). A procedure has been established to separate these 'non-responding' DNAs from 'genuine' DNAs.

The second part of this section looks at the imputation methodology used when new ethnicity questions were introduced in 2011.

The final part reports on work conducted to examine the extent to which the use of imputed data on the LFS leads to estimates which depress estimates of change.

12.1 IMPUTATION AND NON-RESPONDING DNAS

When running LFS tables the DNA ('Does Not Apply') category may be unexpectedly large. This is because certain questions are not asked every quarter (see list overleaf) and some respondents are not contacted in successive waves.

If respondents from one quarter are non-respondents in a subsequent quarter (for wave 2 to wave 5 interviews) then data is carried forward from previous quarters. However, if the question was not asked in the previous quarter there is no data to bring forward, so the response to the question is coded as DNA. As no current data is available for these non-respondents, one way to treat them would be to leave them in the population distribution as effectively "Not known" in the same way as the "Not answered" category is used. Alternatively, if the best estimates for the whole population are required, then by assuming that these cases with missing data have the same distributions as the respondents, they can be eliminated from the survey estimates. To achieve this, an additional weight is required.

To check whether there are non-responding DNAs the variable concerned should be cross tabulated with a variable called IOUTCOME to differentiate between 'genuine' DNAs and non-responding DNAs, which will be shown in code 6 of IOUTCOME (data brought forward from the previous quarter).

The process of imputing the non-responding DNAs is as follows:

(i) calculate: $$\frac{\text{Valid response total}}{(\text{Valid response total}) - (\text{non-responding DNAs})}$$

(ii) multiply each of the valid responses by this factor (exclude DNA)

Example of imputation of non-responding DNAs

The LFS questionnaire explains which groups of people should be asked each question. For example, in the case of UNION the people asked the question consist of those who are:

IF   WRKING=1   did paid work in the reference week
OR JBAWAY=1   temporarily away from a job in reference week
OR TYPSCH12=1,2,3,5,8 working/temporarily away from an employer/
                                    working for a voluntary organisation//environmental
                                    taskforce/ receiving help setting up as self-employed
OR TYPSCH12=9 AND   on a project providing work experience /practical   training
        YTETJB=1   had a paid job in addition to scheme
OR YTETJB=1   paid job in addition to scheme

By filtering on these groups it is possible to produce the following table of UNION by IOUTCOME:

OD (Q4) 2023

Union	Base	Personal response	Proxy response	Data brought forward
Base	32,931,861	15,718,011	10,597,862	6,615,988
Yes	5,337,425	3,421,683	1,915,742	0
No	20,674,539	12,238,228	8,436,311	0
NA	303,909	58,100	245,809	
DNA	6,615,988	0	0	6,615,988

As we would expect, having filtered on only those groups that are actually asked the question, the only DNAs that are picked up are those where the data has been brought forward due to non contact. This can be used as a check to see that no-one else (who should not be asked the question) is being inadvertently asked the question.

It is quite simple to calculate the weight required to adjust the estimates of the non-missing categories and eliminate the non-responding DNAs. The factor is:

32,931,861 / (32,931,861– 6,615,988) = 2.02

This weight can then be used to multiply the frequencies of the valid codes as follows.

UNION
Base   32,931,861
Yes   5,337425 x 2.01839405 =

6,805,715

No   20,674,539 x 2.01839405 =
24,934,844

NA   303,909 x 2.01839405 =
337,396

DNA   0

Variables that may be, or may have been, affected by non-responding DNAs are:

ACCDAY4	EDINS11	MATLVE	SMOCCT	TRONJB	WCHJB
ACCURH(1-4)	FEEIR(1-5)	NOBACK9	SMSOC10	TRVDRV	WCHJB3
AWARE	GOBCK9	NOCUST	SMSOC101	TRVMTH	
BANK	HOLS	NUMILL	SMSOC103	TRVTME	
BHNOTA	HOLSB	QAPL11	SMSOC104	TUCOV	
BHNOTB	ILCURR	REASOFF9	TFEE10(1-5)	TUPRES	
BHNOTC	ILLWRK	ROAD	TIMECODE	TYPILL	
BHPAID	LANG	SMEARNER	TIMEDAYS	TYPINJ	
BNKH11(01-11)	LANGD1	SMHCOMP	TMEOFF	UNION	
BNKHOLF	LANGD2	SMOCCD	TRHR11	VOCQPL11	

12.2 IMPUTATION OF ETHNICITY IN 2011 AND 2001

Changes were made to the ethnicity (as well as national identity and religion) questions in January 2011 to bring them in line with the census data collection on these topics. In April 2011, further changes were made to the ethnicity questions to bring them in line with the Scottish Census data collection. Details of these ethnicity changes on the LFS can be found in the report linked below: http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/articles-and-reports/2011-changes-to-how-ethnicity-is-asked-on-labour-force-survey.pdf

The new questions introduced in 2011 were asked afresh rather than having data rotated into them (as is common practice with all new questions).Therefore, where there was a non-contact in JM11 after a successful interview in the previous quarter (OD10), data couldn't be rotated forward for one wave as usual. Instead these 'data brought forward' cases (IOUTCOME=6) had their values imputed, by using their OD10 responses and mapping them across the new JM11 questions as best as possible,. Some of the new JM11 questions had no equivalent in OD10 so no data could be imputed.

In 2001 there were also changes made to the ethnicity questions to bring them in line with the 2001 census. Here additional imputation procedures were adopted to ensure the greatest possible number of cases has the new ethnicity information for

the spring 2001 quarter. The previous version of this user guide (from 2001) explains the four stages (Augmentation, Re-coding, Modelling, Imputation) used to try to correct the quality issues for this change

12.3 IMPUTATION AND ESTIMATES OF CHANGE

The practice of imputation in the LFS - rolling forwards information from the previous quarter for non-respondents in the current quarter – can be criticised for depressing measures of change. In order to investigate this issue effectively it is necessary to use linked LFS databases.

Consider two consecutive quarters. Then we have full information for respondents contacted in both quarters. However for cases interviewed in the first quarter who failed to respond in the second, imputed values are substituted by rolling forward their answers from the first quarter. About 3% of cases have their values imputed in this way in each quarter. For some individuals, these imputations will be correct and for others they will be incorrect. If a large proportion is correct, then including them will lead to an improvement in the quality of the current quarter's estimates at the cost of only a small bias in the estimates of change from the previous quarter. Alternatively, if a large proportion of them is incorrect then the quality of both the current estimates and the change estimates will suffer.

Although we cannot be certain what the correct value is for a particular non-respondent, we can look at the speed of change among those who respond in successive quarters and at the answers given by these temporary non-respondents in subsequent quarters. It is also important to compare these temporary non-respondents with current respondents in order to assess whether dropping them from the survey, rather than imputing values for them, would create any larger non-response bias.

If non-respondents whose values are imputed resemble respondents to the survey (in terms of their employment status characteristics and propensity to change this status, for example), then it would be appropriate to weight for these non-respondents on the basis of the values and patterns of change observed amongst the respondent population – i.e. there would be nothing to gain by imputing values for them.

However, if non-respondents and respondents are sufficiently different from each other (on non-demographic factors), then information derived from the respondent population is unlikely to be successful in estimating the characteristics of the whole population.

Empirical Evidence

The rather limited evidence we have on this derives from a study undertaken using data from three quarters in 1992/93. The data examined were the numbers in the main economic status categories (mainly those in employment) for respondents in winter 1992/3, spring 1993 and summer 1993; and for winter 1992/3 and summer

1993 for the separate group who did not respond in spring 1993 (and whose data for this quarter were imputed).

The increase in employment for respondents was from 32,174 in winter 1992/3 (55.2% of respondents) to 32,312 in spring 1993 (55.4%) - an increase of 0.4%. For those non-respondents in spring 1993, who had their values imputed from the previous quarter, there was obviously no (recorded) change – 1290 were recorded as employed in both quarters. As one indicator of the level of real change in this latter group, the answers given in winter 1992/3 and summer 1993 were compared, the implication being that if little change is recorded over the longer 6 month period, then it is unlikely that such changes occurred in the two 3 month periods.

The change in employment for respondents was from 32,174 (in winter) to 32,487 (in summer) - up 0.97%. For those people whose data were imputed in spring, the corresponding figures were 1,290 (66.5% of imputed cases) in winter and 1,289 (66.4%) in summer. That is to say, there was virtually no change - which is, of course, what the imputation process assumes for the previous quarter.

So as there was little change in the numbers in employment between winter and summer for those whose spring data were imputed, and the change between winter and spring and between spring and summer for respondents were of the same orders of magnitude (increases of 0.4% and 0.6% respectively), then it seems unlikely that there were substantial counter-balancing moves between employment and the other states between winter and spring and between spring and summer for those whose spring data were imputed.

Although the effect of the imputation on the change in the sample numbers who were recorded as employed was trivial with these data, there is still an argument for avoiding *any* increased risk of bias. Rolling forward data from the previous quarter is only one method of dealing with non-response. In a situation like this, where non-response means that no data is available for the current quarter, the only realistic alternative to imputation is to rely on population weighting. This assumes implicitly that the characteristics of non-respondents are broadly similar to those of respondents with respect to economic status etc.

For these data, the economic status distribution (in winter) of those interviewed in spring and of those not interviewed and imputed in the spring are different, as shown below:

	Winter characteristics of:	
	those who responded in spring	those whose data were imputed in spring
Employees	47.5%	57.6%
Self-employed	6.7%	7.8%
ILO unemployed	5.8%	8.5%
Inactive	39.0%	25.0%

So, the group who were not interviewed in the spring and whose data were imputed from their winter responses had a substantially higher proportion of economically active individuals than the group who responded to the survey in the spring.

The implication of this finding is that to drop these non-responding cases and to rely solely on the population weighting used on the survey to deal with this type of non-response would lose valuable additional information from the survey and hence would probably reduce the quality of the current survey estimates slightly.

SECTION 13 - CONTINUITY AND DISCONTINUITY ON THE LFS

Minimising the risk and impact of losing LFS continuity

As an annual survey up until 1991, the LFS was principally valuable for the in-depth cross-sectional analyses of the labour market (e.g employment, ILO unemployment, total hours worked etc) which it provided. Since its switch from annual to quarterly frequency in 1992, however, a wide range of users of the LFS have increasingly looked towards the survey as a source of time series as well as cross-sectional data. This change in emphasis in the analytical capability of the LFS has increased users' awareness of, and sensitivity to, loss of continuity.

In addition to the use of the LFS for monitoring changes in the labour market - the survey is used for monitoring changes over time in a number of other aspects of people's behaviour which are of interest in various fields of government policy, for example education and training.

13.1 POSSIBLE CAUSES OF DISCONTINUITIES IN LFS DATA

The central aim of the LFS is to categorise the adult population according to the main categories of - in employment, unemployed and economically inactive - and sub-divisions of these, defined according to the guidelines promulgated by the International Labour Organisation (ILO). These variables are to be regarded as the LFS "core" and accorded particular care in respect of their continuity.

Changes in the administrative arrangements for eligibility for unemployment-related or other social security benefits may have an impact on the LFS measures of employment, unemployment or economic inactivity. For example, the switch from Invalidity to Incapacity Benefit, accompanied by the introduction of a more stringent qualifying medical test, might over a period induce a greater degree of job seeking activity in the labour market. However, such changes can never cause discontinuities in the LFS series, as long as the basis of the survey in terms of the ILO definitions remains constant.

ONS will, subject to resource constraints, investigate the impact of administrative changes on LFS estimates. However, it should be recognised that it will almost always be difficult to disentangle such effects from the impact of the general economic or social factors which affect the LFS measures, and that it may not be

practically possible to generate useful estimates of the impact of such administrative changes on the LFS estimates. For example for the introduction of Job Seeker's Allowance, an hypothetical impact over a six month period on the LFS measure of unemployment of the order of 35,000 (which was predicted to have been the approximate impact on the claimant count) would be undetectable in the context of estimates of quarterly changes in ILO unemployment for which the 95% Confidence Limits are ±58,000.

Discontinuities in LFS series have arisen in the past, or could arise in future, because of the following:

(i)   Definitional changes

Any changes in the underlying definitions, on which the estimates are based, could have an impact. For example: the switch to the current ILO definition of unemployment in 1984; and the inclusion of unpaid family workers among the employed population in 1992. In these cases, statistics have been published by ONS describing the impact of the changes on the LFS estimates.

(ii)   Questionnaire changes

If the questionnaire is changed in order to collect new data or to improve the quality of existing items. Effects may most obviously occur in the time series from an existing question if it is changed, more subtle side-effects may occur in the time series for other, related questions as well as to questions where there are alterations in the routing.

For example: in the summer of 1994, the inclusion of a 13-week job-related training question improved the quality of, but caused a discontinuity in, the existing 4-week job-related training information and more recently changes to the LFS qualifications questions improved the quality of the data collected but also introduced discontinuities. In neither case, does an obvious method of estimating the extent of the discontinuity exist.

(iii)   Data processing effects

A number of processes need to be implemented in order to convert the raw returns from LFS interviews into the published estimates. Major changes in the methods or external data used in these processes may in some, but not all, circumstances cause discontinuities in LFS series.

Data editing: a number of minor improvements in LFS editing procedures were introduced at the interviewer stage and this enabled a greater degree of cross-checking of the validity of data with the respondent. The aim was to increase the quality of the LFS data relating to households and families, and no major discontinuities occurred.

Imputation: a particular feature of the LFS - which has been shown to be beneficial for the quality of the data - is that missing responses for people still

resident in the sampled household are substituted by values carried forward from the responses made for the same person in the previous quarter. Where new questions are introduced, or amendments are made, however, this process may not function and a discontinuity may, potentially, arise because of an increased level of question non-response.

Changes in coding frames or classifications: the introduction of a new standard nomenclature, such as the 2020 Standard Occupational Classification or the 2007 Standard Industrial Classification, can cause disruption to time series Guidance was produced about these changes.

Sample weighting to known population totals: changes to the population controls used in the survey can have an impact on LFS estimates. Most recently was the 2022 reweighting exercise applied to all LFS/APS outputs, which used PAYE RTI-based growth rates due to the absence of population estimates during pandemic, a.s well as the 2024 reweighting exercise on the LFS person level weights using the latest available 2021 Census-based National Population Projections at the time.(see section 10)

Future changes in the methodology for sample weighting are a potential risk to continuity. It may be possible to increase the quality of LFS, though a new procedure may simply reduce the sampling errors of the LFS estimates and a discontinuity in the LFS series may not arise.

Seasonal adjustment: a review of LFS seasonally adjusted data is conducted in the spring of each year and a series of revised estimates is published back.

(iv)   Impact of switch from annual to quarterly LFS design

The revised 1996 LFS Historical Supplement describes and assesses the impact of a number of changes that were made in the LFS design, sampling frame and methodology, when the LFS was switched from annual to quarterly frequency. As far as possible, estimates are made of the magnitude of the discrepancies between both the annual and established quarterly surveys and between the introductory and established quarterly surveys.

13.2 CIRCUMSTANCES IN WHICH THE INTRODUCTION OF DISCONTINUITIES MAY BE JUSTIFIED

The introduction of discontinuities to the LFS time series (see table 13.1) is, usually undesirable because of the potential disruption which may be caused to users of the data. However, there are circumstances where the advantages of making changes which may cause discontinuities over-ride the disadvantages, or there are external factors outside the control of ONS.

One over-arching issue in considering whether a discontinuity might be justified concerns the importance of the series affected. Arguments exist to support the view that virtually all LFS series are "important" to one user or another, but the series from the survey relating to the ILO-defined estimates - employment, ILO unemployment

and economic inactivity - yield the "core" information which defines the primary reason for the existence of the LFS.

While the benefits of changing any LFS procedures or any part of the LFS questionnaire, therefore, need to outweigh the disadvantages of possible discontinuities, the balance is strongly in favour of

Some of the circumstances in which discontinuities may be justified are:

(i)   Quality improvements resulting from change

Examples of cases where the benefits of quality improvement have been seen by users to outweigh the problems caused by discontinuities include:

- the LFS qualifications questions - where refinements to the questions have been designed to monitor the National Targets more precisely.

- the LFS disability questions - where the changes have been designed to bring the LFS estimates closer to the concepts of the Disability Discrimination Act.

- the method of determining family and household structure in the LFS - where changes have been made to harmonise the LFS methodology with that used for other household surveys, and hence to improve the quality, and comparability, of the LFS household and family data.

In each of the three examples quoted in the previous paragraph, ONS worked closely with the Department for Education and Skills to explore the extent of the discontinuities caused and, where possible, to make allowance for them in LFS time series.

(ii)   External factors outside ONS control

A potential external source of impact on the LFS questionnaire, and hence on the continuity of LFS series, is a change in the EU Regulation covering the conduct of the LFS. A new Regulation for a continuous LFS was introduced in 1998 which introduces some changes to the LFS questionnaire requirement.

ONS have consistently pursued a vigorous defence of the existing UK LFS methodology and questionnaire in the discussions of the Eurostat Working Party which led up to the development of the new Regulation. As a result, changes to the existing UK LFS questionnaire needed to conform to the new Regulation will be minimal and will certainly not affect the "core" LFS series.

(iii)   Major survey re-design

For example the up-grade of the LFS from an annual to a quarterly survey**.**

(iv)   Change of contractor

Should a change of contractor occur, there would inevitably be some impact on the continuity of LFS data, even in the core series, resulting from the well-documented "contractor effect".

13.3 STRATEGY FOR CONSIDERATION OF LFS CONTINUITY.

The issues relating to the maintenance of continuous time series from a household survey, such as the LFS, are very complex. No overall prescriptive basis exists for dealing with all the circumstances that may arise, but the guidelines - should be valuable as a basis for future consideration of the issues.
The previous version of this user guide from 2011 contains more information on the strategy and guidelines for minimising the risk and impact of a loss of LFS continuity, below is a summary:

(i) **Attempt to recognise the risks of discontinuity in advance**. If changes to the LFS (e.g. survey method, questionnaire etc) are planned, then part of the planning process should specifically include an assessment of the potential for discontinuities

(ii) **To assess the benefits and disadvantages of changes to the LFS.** Any benefits of changes to the LFS will be assessed along with the impact these changes may have on the continuity of LFS time series and, where possible, decisions on the implementation of such changes will be taken by ONS in consultation with LFS users, in the light of all the relevant factors.

(iii) **To consult with users.** ONS will attempt to resolve discontinuities, subject to resource constraints, in consultation with appropriate LFS users. Input from subject-matter specialists within OGDs will be actively encouraged.

(iv) **To determine the appropriate response to the discontinuity or risk of discontinuity**
   a. By bringing the discontinuity to users' notice for example by indicating the discontinuity in LFS tables and user guides (e.g volume 3 which gives details for each LFS variable)
   b. By revising LFS historical series onto a consistent basis e.g reweighed data, seasonal adjustment, .
   c. By publishing dual estimates for one or more benchmark quarters e.g changes to classifications
   d. By indirectly estimating the size of discontinuity

(v) **To establish a relevant dialogue with other National Statistical Institutes**
   For sharing knowledge about the potential risks to data continuity and means of dealing with such losses of continuity.

13.4 DISCONTINUITIES ON THE LABOUR FORCE SURVEY

Table 13.1: A list of topics where there have been discontinuities on the Labour Force Survey

Topic	Time of discontinuity
Employment	1983 Spring 1992
Unemployment	1984 Spring/Summer 1992 Spring 1993
When left last job Redundancies in the last three months	Spring 1992
When started with current employer	Spring 1992/Summer 1993
Redundancies in the last three Months	Spring 1995
Reasons for economic inactivity	Spring 1992/Summer 1993
Long term health problem	Summer 1993/4 Spring 1996
Qualifications	Spring 1996
Numbers of graduates	1991-1993
Coding of occupations	1991/2001/2011/2021
Coding of industry	Winter 1993-94/2007
Household and family data	Spring 1992
Job-related training	Summer 1994
Ethnic origin and nationality	Spring1992/Spring 2001
Irish nationality	Winter 1994-95 Autumn 1995
Temporary employees	Spring 1992
Northern Ireland qualifications	Spring 1996
Education courses	Spring 1997
Disability data	Spring 1997
Benefits questions	October 1999
Sickness absence	October 1999
Enhancement reference period	Spring 2004
Number of O-level/GCSE etc passes held	Spring 2004

Details of these discontinuities before 2004 can be found in the 2011 version of this user guide.

A summary of the changes from 2004 can be found below, the LFS user guide volume 3 (details of the LFS variables) is a useful place to look and will provide more details particularly on the variables:

Enhancement survey questions reference period amended from Spring 2004
From March to May, the reference period used for respondents partaking in the enhancements, changed from three years to one year. Thus, respondents are now asked questions with reference to their situation within the last year instead of the last three years.

Education questions regarding number of O-level, GCSE etc passes already held from Spring 2005
Prior to Spring 2004, the LFS provided 3 broad categories to respondents in terms of how many O-level, or GCSE etc passes that they held. From Spring 2004, as requested from the Department for Education and Skills, respondents were given a more detailed set of response options with the hopes of gaining more specific data. It was found that this produced a larger than expected increase in the number of don't know responses, due to respondents not being able to provide the exact number of passes that they had. The end result showed that there was an increase of about 2% in each category, which could only be explained due to the change in the question. As a result, from Spring 2005, DfES requested that the LFS return to the response options used prior to Spring 2004. There are now follow on questions in place in order to gain the more specific detail.

Type of agreed work agreements
Between the first quarter of 2004 and JS13, there was a check in the questionnaire that did not allow respondents to say that they worked shift work at SHFTWK99 and then go on to say that they worked zero hours at FLEX10. However, as SHFTWK99 is only asked in wave 1 and in AJ, this check only affected AJ respondents and all wave 1 respondents. From JS13 onwards, this check has been removed i.e. zero hours contracts and shift work are no longer deemed incompatible. Analysts should be aware of this when doing any analysis surrounding zero hours contracts and are advised to restrict their analysis to OD quarters. Please note that this does not affect data sets prior to 2004 and after AJ13.

Standard Industrial Classification
The industry class to which people in employment are coded in the LFS switched to SIC 2007 in January 2009. From then on, all cases were assigned an industry code on the new basis. This included respondents who had been surveyed in the previous quarter whose employment situation had not changed. There was no dual coding.

The transition to the new classification was accompanied by the implementation of a new automatic coding tool for LFS interviewers. A similar tool had been in place for the coding of occupations to the Standard Occupational Classification for several years, but prior to 2009, industry had been coded manually using a printed volume. The new tool is seen as a significant improvement in the coding of industries in the LFS. Its introduction has brought greater consistency since cases with the same description are more likely to be allocated the same code with the coding tool than with the previous approach.

Users should be aware that, for the quarterly time series of employment on a SIC 2007 basis, there are several step changes at Section and Division level between Q4 2008 and Q1 2009. Investigative analysis has shown that these were caused

primarily by the introduction of the new coding tool. Some limitations in the mapping between the two classifications also contributed, but to a much lesser extent.

Some of the main step changes at Q1 2009 when deriving time series on a SIC 2007 basis can be explained further as follows:
• Sections M and N (Professional, Scientific and Technical activities and Administrative & Support Services) gained significantly from people formerly coded to: Manufacturing; Construction; Information and Communication (J); Recreational (R); and Activities of households as employers.
• Manufacturing (Section C) has decreased in size, in particular losing people to various services within Sections M and N: and also to Distribution (G).
• Education (Section P) has gained mostly from people formerly coded to: Public Admin (O); Arts, Entertainment & Recreation (R) and Health & Social Work (Q).
• Section T (Households as Employers) reduced by 55%, losing people mostly to Admin & Support Services (Section N).
More details can be found in the LFS user guide volume 3.

Disability
Discontinuity was identified in disability rates reported in the LFS between quarter four 2009 and quarter one 2010.
Analysis generally showed that the characteristics of those who were disabled (and those already in the survey who 'became' disabled) did not change significantly over time. The few characteristics that might have been associated with the initial increase in disability were found in subsequent quarters not to be associated with the *sustained* high levels of disability and were therefore thought to be random.

Examination of the survey design, questionnaire wording and routing, and anecdotal evidence from interviewers, did not reveal any significant changes over time. The only change to the administration of the questionnaire was the addition of a short introduction at the start of the disability module:
"I should now like to ask you a few questions about your health. These questions will help us estimate the number of people in the country who have health problems."

The impact of this introduction is thought to be positive in that it prepares respondents for the set of disability questions. Any increase caused by this change should result in a more complete measure of actual disability. This change to the introduction added to the disability module in quarter one 2010 is thought to be the key driver of the step increase in disability. The earlier estimates can still be considered 'best estimates' for those periods and should give a robust picture of changes over time, however, direct comparisons between pre- and post-Q1-10 estimates should not be made.

State Pension Age
Between April 2010 and April 2020, women's State pension age in the UK will increase from 60 to 65, at the rate of one month every two months. From August 2010, ONS publish headline employment and inactivity rates based on the 16-64 population. No change was made concerning the unemployment rate, which was, and is still based on the population aged 16 and over. The LFS definition of pension age has changed from 65 plus for men and 60 plus for women to 65 plus for men

and women; all children under 16 years are classified as economically inactive. Therefore, from April 2010 any routing that previously used the population 'males 16-64 and females 16-59' should change to 'all those aged 16-64', and routing that previously used the population 'males 65+ and females 60+ (pension age)' should change to 'all those aged 65+'.

Sickness
In AJ10 some amendments were made to the LFS questionnaire in order to improve the collection of data on days taken off due to sickness absence. Prior to AJ10, respondents who were employed but had been off for the whole reference week (or longer) were stating at ACTWKDY that they were not working and so didn't get asked ILLWK or ILLDAYS1-7. A check was introduced at the question ACTWKDY with the purpose of increasing the number of people either self-employed or off work for reasons of sickness or injury disclosing their scheduled work days. Despite improving the accuracy of the data collected the required changes resulted in a discontinuity in the time series. The introduction of the check has led to an increase in people reporting five days or more sickness absence who were previously being missed.

Standard Occupation Classification
In the development of SOC2010, there is a significant decrease in the numbers coded to the Managers and Senior Officials major group. The increase in the Professional Occupations major group is likely to be the corollary of this decrease.

National Identity and Religion
Changes were made to the national identity and religion questions in January 2011 to bring them in line with the census. Since these questions were different to those asked in OD10 they were asked afresh and data wasn't rotated into them (as is common practice with all new questions). Values for the data brought forward cases in JM11 (cases that responded in OD10 but were non-contacts in JM11i.e had an IOUCTOME of 6) were imputed using their OD10 responses and mapping them across the new JM11 questions as best as possible. Some of the new JM11 questions have no equivalent in OD11 so no data could be imputed.

Ethnicity
Changes were made to the ethnicity questions in January 2011 to bring them in line with the census. In April 2011, further changes were made to the ethnicity questions to bring them in line with the Scottish Census data collection.
The new Ethnicity questions introduced in 2011 were asked afresh rather than having data rotated into them Values for the data brought forward cases in JM11 (cases that responded in OD10 but were non-contacts in JM11i.e had an IOUCTOME of 6) were imputed using their OD10 responses and mapping them across the new JM11 questions as best as possible. Some of the new JM11 questions have no equivalent in OD11 so no data could be imputed.
More information about the changes to ethnicity can be found in this paper: http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/articles-and-reports/2011-changes-to-how-ethnicity-is-asked-on-labour-force-survey.pdf

Education
Due to an error in the routing to the education section in the JM11 and AJ11 questionnaires, the education section which should have been asked of those aged 16-69 or older and in employment was only asked of those aged 16-64 or older and in employment. 65-69 year olds who should have been asked these questions were not and there will be more missing education data in JM11, AJ11 (and some impact on JS11 for DBF cases). This issue should stop impacting the data from OD11.

Also from JM11 more information on foreign qualifications that are recognised in the UK was collected; this can be used to assign more accurate levels of highest qualification. Therefore a large number of respondents who previously ended up as 'other' on the HIQUAL and LEVQUAL DV's are now assigned to appropriate qualifications/levels. This means that there is a clear break in the education time series as we see the numbers classified as 'other' dropping and from JM11 and certain qualification/levels absorbing these cases which are now assigned to a qualification/level. This should improve the accuracy of the data but affects the comparability of qualifications over time.

Disability
From JM12, the method calculating the group '4 – not disabled' changed with respondents who answered LNGLIM=1 or 2 not being classed as in category 4 like previously and instead given a value of -9 for DISCURR, this will therefore create a discontinuity in the data.

In OD13, the variable DISCURR13 was introduced, however the DDA disabled (current disability) category within DISCURR13 is not the most appropriate one to use. It is not comparable to the corresponding category in DISCURR, prior to AJ13, because of changes to the questions, and it no longer measures the DDA definition of disability. Neither does it measure the latest Equality Act definition of disability, which is available from the DISEA variable instead.

Benefits
In AJ14 to OD14 there was no CLAIMS variable on the dataset, so there is a break in the series. CLAIMS14 was introduced in JM15. Other recent benefit changes include from AJ13 Universal Credit and Personal Independence Payment was introduced and Council Tax reduction replaced Council Tax Benefits from JS13.

Education
From JM15 new Scottish qualifications have been added to the education questions, which also affect the derived variables for example HIQUAL11 became HIQUAL15. Care needs to be taken when using the variables GCSE4, QGCSE4 and GCSEFUL, as the variable name wasn't changed even though the categories did, previously category 6 was "none of these" but from 2015 category 6 is "Scottish Nationals level 5" and category 7 is "none of these".

Imputation of personal characteristics on LFS Household datasets
A review of the imputation methods used in LFS Household and Family analysis resulted in a change from JM15 onwards. It was decided that it wasn't appropriate to

impute any personal characteristic variables (e.g. Religion, Ethnicity, Country of Birth, Nationality, National Identity etc) using the LFS donor imputation method. This method is primarily focused to ensure the 'economic status' of all individuals within a household is known, allowing analysis of the combined economic status of households.

This means that from 2015 there will be larger amounts of 'missing's' (-8's/-9's) for these personal characteristic variables than before. Therefore if you need to carry out any time series analysis of Households/Families which also includes personal characteristic variables covering this time period, then it is advised to filter off ioutcome=3 cases from all periods to remove this inconsistent treatment of non-responders.

SECTION 14 – QUALITY

The LFS Performance & Quality Monitoring Report (PQM)

The LFS reports quality issues in the LFS Performance and Quality Monitoring (PQM) Report, which is published quarterly on the ONS website.: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentand employeetypes/methodologies/labourforcesurveyperformanceandqualitymonitoringre ports

The LFS PQM was first produced in 1992 and contains detailed information on aspects of survey quality, including achieved sample size and response rates. The PQM was revised in 2006 to coincide with the first release of LFS data on a calendar quarter basis beginning with the January - March 2006 quarter. There were a number of reasons for revising the PQM at that time. The first was that the format of the PQM had remained largely unchanged since its launch in the early 1990s and its contents were considered to be in need of review. A number of key indicators of data quality were identified as missing from the report, and some of the items being reported on were no longer thought to be of interest. The process of identifying and agreeing changes to the PQM was done in consultation with customers (internal and external) to ensure that it met their requirements. In addition, the format and content of the LFS PQM was brought it into line with the corporate approach to quality reporting which the ONS implemented with the support of the Statistics Commission.

PQM Contents

Overall, the PQM has adopted the ONS corporate approach to quality reporting in which quality is reported against the six European Statistical System (ESS) dimensions of quality, namely: relevance; accuracy; timeliness & punctuality; accessibility and clarity; comparability; and coherence. Each dimension of quality is defined in the PQM.

The PQM contains a 2-page 'Executive Summary' that briefly describes the current status of the survey in terms of:

- achieved sample size
- response rates
- delivery dates of data
- quarter to quarter changes, including any changes to the survey, such as new or amended questions, and
- fieldwork issues, particularly those which are likely to have an impact on data quality.

Specific quality issues reported in detail in the PQM include:

- the relevance of the LFS including its primary purpose, users and uses, strengths and limitations, and key definitions.
- The accuracy of the LFS including
  - a time series of achieved number of household and person interviews for GB & UK.
  - sampling variability estimates for a number of key LFS variables for the UK
  - wave-specific response rates and an overall response rate for the quarter for both GB & UK.
  - a time series of wave-specific response rates for GB
  - the composition of non-response and how it has changed over time
  - wave-specific response rates by Government Office Region for the quarter
  - proxy response rates for the quarter
  - income response rates by NS-SEC for the quarter, and
  - attrition rates by key person level characteristics for the quarter
- information on timeliness and punctuality including delivery dates of data
- information on accessibility and clarity, including various access points
- information on comparability, including definitions, quarter to quarter changes and fieldwork issues
- information on coherence with other sources of data on the labour market, including the strengths and limitations of the LFS
- a summary of methods used in the LFS
- technical definitions, and
- website references

PQM Publication Date

The PQM is released to coincide with the release of the quarterly LFS data to which it relates. As the PQM contains a small amount of market sensitive data in the form of estimates for key variables, it is released in accordance with the published timetable for the Labour Market Statistics Integrated First Release. This is a requirement of the National Statistics Code of Practice and Protocol on Release Practices.

Other Quality Reports relevant to the LFS

In addition to the PQM, the LFS also has a Summary Quality Report (SQR) which also provides users with information on fitness for purpose of the LFS, and contains qualitative information covering the six ESS dimensions of quality covered by the PQM, and a summary of methods used to compile the output. What it does not contain, however, is quantitative information relevant to each quarterly release of LFS data, as this is contained in the PQM. In addition to the LFS SQR, quality issues relevant to the LFS can be found in the Labour Market Statistics (LMS) SQR, along with quality issues relating to the other components of labour market statistics. The LFS and LMS SQRs are published on the NS website and are updated only when there is a change to the qualitative information they contain.

SECTION 15 – HARMONISATION

From 2017, there has been a change from 'Primary Principles' and 'Secondary Principles' to 'GSS Harmonised Principles'
The current information on harmonisation can be found here:
https://gss.civilservice.gov.uk/guidances/harmonisation/
The section below will be updated in the next update of this user guide.

Background

The United Kingdom conducts a wide range of Government surveys of persons and households, which provide sources of social and economic statistics. These surveys were designed at different times, to meet different needs, and have been commissioned by a range of departments. Consequently, the surveys were developed to a significant degree in isolation from each other. This resulted in a lack of cohesion, with differences arising in concepts and definitions, in design, in fieldwork and processing practices and in outputs.

In an attempt to overcome these shortcomings the Social Survey Division of the (then) Office of Population Censuses and Surveys agreed to undertake work to introduce common classifications, definitions and standards for social survey questions, and to improve comparability between social statistics. All this with a view to 'harmonising' the surveys as far as possible without compromising or jeopardising their objectives, and to provide a robust methodological structure within which future developments to these surveys could be framed.

Harmonisation concentrated initially on standardising the inputs to surveys and a differentiation was established between 'primary' questions and concepts (i.e. relevant to all surveys) and those of a 'secondary' nature (relevant to a subset of surveys).

ONS is increasingly positioning itself to make greater use of administrative data, for example, to support Neighbourhood Statistics. In order to exploit such data it will be important to extend the principles of harmonisation beyond surveys and the Census. This is likely to be the focus of new harmonisation activities over the next few years. Other factors influencing harmonisation activities include:

Eurostat requirements – developing harmonised Key Social Indicators
Emerging social topics such as social capital, e-society and cultural identity.

Harmonisation and the LFS - Inputs - Potential for Discontinuities

Whilst the benefits of harmonisation are clear, it is also the case that changing questions or interviewing practices risks the possible introduction of discontinuities. For example, analysis of responses to the harmonised questions on ethnic origin and housing tenure questions in 1996 and further changes ti ethnic origins categories in 1997 showed that they had not caused significant discontinuities.. From Spring 2001, the Labour Force Survey introduced new questions on ethnicity based on recommended output classification of ethnic groups from the 2001 Census. No comparison should be made between the old and new ethnic classifications in the

LFS, because not only are the categories different but, the questions and coding of answers underlying the data are also very different. A similar change was made to ethnicity in 2011 to bring the LFS in line with the census further information on this change can be found http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/articles-and-reports/2011-changes-to-how-ethnicity-is-asked-on-labour-force-survey.pdf

Harmonisation and the LFS - Outputs

The harmonised output categories for economic status are consistent with those used in the LFS. Their development resulted in an improvement in the routing of the harmonised question on reasons for economic inactivity to allow the classification of all such persons, consistently with the LFS.

Since Spring 1996, some harmonised questions have been incorporated in the main LFS, as not all of the topics are relevant to the LFS, as seen in Table 15.1.

Table 15.1 Harmonised Questions for Government Social Surveys - LFS (as at December 2015)

Variable	Whether harmonised in the LFS	Comments
1. Primary set		
Household response unit	No	See (i)
Gender	Yes	
Date of birth	Yes	
Age	Yes	
Legal marital status	Yes	
Living arrangements	Yes	
Who owns or rents accommodation/Tenure	Yes	
Household Reference Person (HRP)	Yes	
Relationship to HRP	Yes	
Ethnic origin	Yes	
National Identity	Yes	
Economic status	No	See (ii)
Employment status	No	See (ii)
Industry – SIC code	No	See (ii)
Occupation – SOC code	No	See (ii)
National Statistics Socio-economic Classification (NS-SEC)	No	See (ii)
Socio-economic group	No	See (ii)
Full-time/part-time work	No	See (ii)
General Health	Yes	
Long lasting health conditions and illnesses	Yes	
Impairments	No	
Activity restriction	Yes	
Geography – use of GORs	Yes	
2. Secondary set		
Benefits and tax credits	No	
Income	No	
Housing costs		
Selected job details	No	See (ii)
Accommodation type of household	No	
Length of residence	Yes	
National Identity	Yes	
Religion	No	See (iii)
Internet Access	No	
Sexual Identity	Yes	
Qualifications	Yes	
Educational attainment	No	Though can be derived from input variable (iv)
Personal Well-being	Yes	

The following primary topics are not included on the LFS : Agency workers, Carers and Mobile Phone use. The following secondary topics are not included on the LFS: Consumer durables, Household motor vehicles, Rooms available to household, Housing costs and benefits, Crime and fear of crime,

Notes:

(i) The harmonised definition of the household response unit is "one person living alone or a group of people (not necessarily related) living at the same address who share cooking facilities and share a living room or sitting room or dining area.   The LFS definition differs slightly. The LFS adds students who live in halls of residence in term-time and residents in National Health Service accommodation to the coverage allowed in the harmonised definition, but these are clearly identified and the harmonised definition can be derived.

(ii) The LFS uses the International Labour Office (ILO) definition of economic status. The harmonised input is *based* on this definition, but differs in minor respects since the ILO standards depend on more complex questions than are possible for a harmonised question for general social surveys. The minor differences with the LFS affect people who were on a government supported training scheme and the classification between full-time and part-time work. This departure from the harmonised question affects several topics, including economic activity and usual hours in main job. The result of this is while the questions relating to economic activity are mostly harmonised; the outputs differ from the
harmonised outputs.

(iii) The religion question on the Labour Force Survey for Northern Ireland asks about religious denomination, with the main difference being nine categories, where 'other protestant' and 'other religion' are combined compared to sixteen on the harmonised principle question.

(iv) In 2004, ONS developed a set of simple questions which would measure educational attainment through the **highest qualification** obtained. However, the LFS has a substantive interest in qualifications and have a set of questions which collect all qualifications, the highest qualification in a scale of the analysts choosing can then be derived.

SECTION 16 - USES OF THE LFS

Introduction

The Labour Force Survey (LFS) began as a condition of UK membership of the European Community and was carried out biennially from 1973 - 1981 and annually from 1984 - 1991. Over this time Government departments, especially the Employment Department, found the information collected in the LFS increasingly valuable in the framing of social and economic policy. In 1990, the Secretary of State for Employment announced the development of a quarterly LFS which began in spring 1992.

The main purpose of the quarterly LFS is to provide information needed to develop, manage, evaluate and report on labour market policies. Currently, interviewing takes place in approximately 37,000 households a quarter, yielding labour market and demographic information about some 90,000 individuals. Its main strengths are that it provides a self-contained, integrated source of information about the Labour market activity (or inactivity) of the whole (household) population, based on a large sample size, and that it uses the internationally standard definitions of employment and unemployment recommended by the International Labour Organisation (ILO).

Topics covered by the LFS

The LFS provides regular information relating to the following topics:

- demographic characteristics of the population;
- employment, unemployment and inactivity;
- qualifications held and in the process of being attained;
- job-related training;
- trade union membership and the coverage of collective bargaining,
- industrial accidents and their causes;
- work related illnesses;
- earnings and sources of income

Macro-economic monitoring

The quarterly LFS is highly valuable in helping to assess changes in the labour market. First key results are now published one and a half months after the survey period ends, with full results available two months later. Main indicators regularly published from the LFS include -

- ILO unemployment, total employment, ILO unemployment rate and economic activity rate (employment and unemployment as a percentage of the total population), by age group;

- employees and self-employed people, full- and part-time workers, second jobs and temporary workers, by industry and occupation;

- average actual working hours and total hours worked in the economy;

- redundancies;

- reasons why people are economically inactive (not employed or unemployed) and whether they would like to work, including groups such as:
     - discouraged workers - those who say they would like to work but have not looked for work recently because they believe no jobs are available and therefore are excluded from measures of unemployment;
     - people (usually women) looking after the family or home;
     - students;
     - retired people;
     - people unable to work because they are sick or disabled.

The LFS is useful as an alternative source of information, relying, on a different collection method, with which to compare the trends shown by the claimant count of unemployment and the surveys of employers about employees. Each source has its own strengths and weaknesses[28] in particular, the articulated nature of the LFS means that it can provide important information to explain such unexpected (to the casual observer) phenomena as a fall in unemployment at the same time as a fall, or a smaller rise, in employment. The LFS may be able to show that the difference is explainable, for instance, by an increase in the number of people in full-time education, information which is not available from unemployment or employment records. The LFS also provides estimates for sections of the labour force who are not covered by the employer surveys, such as the self-employed and temporary employees, or the claimant count of unemployment such as those ineligible for unemployment-related benefits (e.g. most under 15 year olds), and those with a low propensity to claim (such as married women).

The LFS provides the basis for labour force projections which provide an assessment of the likely chances in the composition of the labour force over the next 10-15 years. These projections assist in the formulation of policies which will take account of predictable chances in the economically active population.

The "flexible" labour market

The LFS collects a wide range of information about people's employment, and is the only source of quarterly statistics on self-employment, temporary workers and the type of contract they have - fixed period/task, agency work, casual etc, none of which is available as frequently from any other source. Because it is a survey of people not employers, the LFS can show the mix of employment types varying from full-time to part-time and temporary, self-employed and unpaid working for a family business. This basic information can be linked to more in depth results such as the reasons why people work part-time, such as the proportion who do so because they could not get a full-time job. The LFS is also the only regular source of estimates of the extent of home-working.

The survey collects information on usual and actual working hours, including separate figures for overtime, used to show, for example, that this country has the

---

[28] For a comparison between LFS and claimant count estimates of unemployment see *Labour Market Trends,* February 2004.

most varied pattern of working hours in Europe. Questions are also asked about evening, Saturday and Sunday work. A research feature in the January 2000 *Labour Market Trends* explored the data available from the LFS on working patterns, and describes the characteristics of people who work flexibly.

Another aspect of the flexibility of the labour market is labour mobility and the LFS helps to monitor this by means of questions asking people about the job they were doing one year earlier, and whether they moved to find work. The survey also asks how long employees have been working with their current employer and if they have left a job recently, people are asked the reason why. It is also possible to identify people who have returned to the labour market since the previous year, such as women returning to work after a break to bring up a family.

Regional statistics

Regional data have always been available from the LFS and now a limited number of key variables on employment and training are provided for local authority districts and Training and Enterprise Council (TEC) areas. This helps TECs and the Government Offices for the regions to assess local labour markets to inform their planning processes and to advise local people and businesses. Although small area data are not the LFS's strong point, the estimates which are available go some way towards meeting the need for information about areas such as inner cities and rural areas, whose special needs are considered on an interdepartmental basis.

The characteristics of the unemployed

The information about the characteristics of unemployed people which is available from the LFS, such as marital status and qualifications, complements the information collected about benefit claimants. The LFS is able to identify groups of interest such as disabled people and lone parents who may face particular problems in getting work, and people from ethnic minorities. It also provides information about the duration of unemployment, and the occupations and industries where the unemployed previously worked.

The Department for Work and Pensions (DWP) uses information from the LFS to help devise and assess services to help people not in work. They are interested in the reasons why people do or do not seek work, and the methods they use, both to judge the effectiveness of their policies and to encourage active and effective job search. The LFS is the main source for monitoring redundancies. An article "Redundancies in the UK", *Labour Market Trends,* May 2004 describes the characteristics of redundant workers and this information also helps the DWP to improve their understanding of this group and the influences on their chances of returning to work.

The LFS uses the internationally standard ILO definition of unemployment[29]. Respondents are also asked whether they were claiming unemployment related

---

[29] The ILO definition of unemployment covers persons: without a job, available to start work in the next fortnight and had actively looked for work in the last four weeks or had found a job and were waiting to start.

benefits. The LFS helped to inform Government about the number of people who were likely to be affected by the changes from Invalidity Benefit and Sickness Benefit to Incapacity Benefit (introduced in April 1995) and from Unemployment Benefit to the Jobseeker's Allowance in 1996. It helped to monitor the effects of the introduction of Incapacity Benefit and JSA both on the claimant count of unemployment and on the ILO measure from the LFS.

Training and qualifications

A number of the Department for Education' (DfE) publications, including Trends in Education and Skills, make extensive use of the LFS. The survey is a key source of information about the amount and type of training done (particularly job related training). Information on the qualifications and employment status achieved by people in different categories - women and ethnic minority groups, in particular - helps to inform policy on further action in the area of training which may be beneficial in promoting equality of opportunity in the labour market.

The Skills Funding Agency (formerly known as the Learning and Skills Council) is responsible for planning and funding vocational education and training in England and it uses the Labour Force Survey as a major source of information when evaluating their effectiveness. Such information also forms part of a range of indicators used to assess the effectiveness of the Training and Enterprise Councils contracted to manage the provision of training for young people and unemployed adults around the country.

Work relating to policies and programmes aimed at increasing adult commitment to learning, (eg, Career Development Loans, Small Firms Training Loans) requires information from the LFS as comparative background information when monitoring the performance of such programmes in terms of participation rates of groups including women, people with disabilities and those from ethnic minorities.

The youth labour market

The LFS is an important source of information about the youth labour market. In particular, it provides up-to-date, quarterly, information about whether young people are in education, which can be combined with information about their economic activity to reflect the multiple activities that they are often engaged in. The LFS is also the primary source of statistics on apprenticeships.

Working conditions

The LFS helps to monitor the coverage of the provisions of the employment protection legislation and to assess the number of people who might be affected by proposed changes. The survey provides estimates of the numbers of employees who qualify for the right to go to an Employment Tribunal if they feel they have been dismissed unfairly (i.e. having completed one years' service). This helps to forecast the number of cases likely to come to the Employment Tribunals. The LFS also provides information on the number of people in small workplaces, where legislation may create a different burden. The LFS also is the only regular source of information

on the holiday entitlements of full- and part-time employees which is of interest in relation to the EU directive on working time.

Trade union membership

The LFS is an important source of information about the level of trade union membership, filling gaps in other sources. The demographic and employment data collected by the LFS is useful in analysing the extent of trade union membership among different groups in the population (e.g. ethnic minorities), sectors of industry, small workplaces, the public sector etc. The LFS also provides a measure of the extent to which employees' pay and conditions are determined by collective bargaining arrangements. These data provide a useful adjunct to workplace-based estimates of collective bargaining[30]. An article used to appear every year in *Labour Market Trends* on this topic.

Incomes

Since winter 1992/93, the LFS in Great Britain has included questions on employees' earnings and other household income. After careful evaluation, these data were released for public use in December 1994 and described in an article in *Employment Gazette.* Income questions were included in the LFS in Northern Ireland from Winter 1994/5. There are other sources of earnings data (e.g. the Annual Survey of Hours and Earnings (ASHE)), but the LFS data is largely unique in that it covers groups such as temporary employees, part-timers and the low-paid, who are not necessarily covered by employers' records. For this reason, the LFS is a key source of data for the Low Pay Commission when setting the National Minimum Wage[31]. The LFS has been used extensively to explore the relationship between pay and qualifications[32]. It has also been used to provide data for the European Union survey on the Structure of Earnings.

Equal opportunities at work for women, people from ethnic minorities, people with disabilities and older workers.

The LFS is a key source of statistics on the characteristics and labour market status of people from different ethnic groups, women, people with health problems and disabilities and older workers. This information is used in monitoring and promoting equal opportunities regardless of race, sex, disability or age, both in the workplace and in other fields covered by government.

The information available from the LFS assists in taking into account relevant factors such as levels of qualification and age when considering the position of particular groups in the labour force, and possible reasons for differences in employment and unemployment levels between them. For example, LFS results contribute to the monitoring of the industrial and occupational segregation of ethnic minority people

---

[30] For example, see Inside the Workplace: First Findings from the 2004 Workplace Employment Relations Survey, DTI

[31] The Annual Report of the Low Pay Commission includes extensive analysis based on LFS earnings data ()

[32] For example, Walker and Zhu, "Education, Earnings and Productivity: recent UK evidence", *Labour Market Trends,* March 2003.

and women, and their progress in achieving managerial positions. The survey also provides information about the types of work done by people with health problems and disabilities, and the types of work previously done by those who have given up employment.

The Equal Human Rights Commission (formerly known as the Equal Opportunities Commission) makes great use of the LFS, including many items based on the survey, in their annual publication *Facts about Women and Men in Britain.*

Households and families

The LFS records information about all members of a household so it is possible to look at family and household characteristics. This aspect of the data has most commonly been used to monitor labour market participation in households. Information about women with dependent children, including lone mothers, is available from the LFS and is used to monitor their participation in the labour market, and to help assess the support needed by working mothers, through childcare provision and other policies. There is also interest in questions such as whether unemployed people tend to have unemployed partners and whether people on low earnings are often in low-income households. Rather more work on employment and earnings at the household or family level has been done by academic researchers, using the LFS over a 10-year period, for example[33]. Data from the LFS on incomes has contributed to the debate on low-income households and the concept of a minimum wage.

Work-related accidents and illness

The Health and Safety Executive (HSE) requires a benchmark against which to interpret the information on workplace accidents reported by employers, which is known to be incomplete. They are interested to know both the level and trends in workplace accidents and the variation in risks between the main sectors of industry. The LFS results were a major input to a recent review of the reporting regulations and will be used to judge whether or not the revised regulations are working. Data are used to inform the allocation of inspectors based on the level of risks at a detailed industry level, and the Annual Report to the Health and Safety Commission regularly features data from the LFS.

Longitudinal analysis

The design of the LFS makes it possible to conduct longitudinal analysis. Datasets are produced linking respectively two and five consecutive waves of data, including all people of working age who respond at each of the waves (see user guide volume 11 for more information). To date, analyses of these data has been largely restricted to the study of labour market flows[34]. There is also a two-year APS longitudinal dataset which links together two January-December APS periods

---

[33] For example, Harrop and Moss, "Working parents: trends in the 1980's", *Employment Gazette,* October 1994.

[34] For example, "People leaving economic inactivity: characteristics and flows", *Labour Market Trends*, April 2002.

Links with International Organisations

The LFS was a European Union survey. The UK LFS included all the questions required by the EU and the data was sent to Eurostat (the Statistical Office of the European Communities) each quarter. The ILO recommended definitions of employment and unemployment, which are used in the LFS are also used in similar surveys in other major nations of the world such as the USA, Canada and Australia. LFS data are also widely used by organisations such as the Council of Europe, the United Nations, ILO and OECD for international comparisons.

LFS data was used to work out the cost to the UK of various proposed EU directives relating to employment conditions. Information about foreign nationals living and working in the UK and about corporate transfers required by the European Union to monitor the freedom of movement of workers within the EU is obtained through the LFS. The LFS also asks about people who have worked, or applied for a job abroad in the last five years.

Other Government Departments and Agencies

The LFS is widely used by Government Departments for analysis of the labour market and to develop government policies in this field. Many also use the LFS for purposes not directly related to the labour market. Some Departments sponsor a limited number of questions in the survey. For example, the Department for Transport sponsor questions on place of work, mode of travel and time taken to travel to work to supplement those collected in the decennial population census. The Home Office makes significant use of the questions on ethnicity and country of birth asked in the LFS to support policy on race and immigration issues.

Other Departments who do not sponsor questions in the LFS, use data collected in the survey primarily for different purposes. The LFS is the only statistical source of information between decennial population censuses which gives estimates of the size of the different ethnic minority populations in Great Britain and this information is used by ONS. ONS also uses the LFS to obtain estimates of the numbers, and characteristics, of households and families, especially of one-parent families. The Home Office uses estimates from the LFS as a benchmark against which to monitor different ethnic groups in the criminal justice system. The Department for Education uses the LFS for assessing the educational participation and qualifications of the population as a whole. HM Treasury is interested in the potential of the LFS to provide information of labour inputs (as hours worked) for calculations of industrial productivity. The Monetary Policy Committee who advise the Treasury on the economy, and the Bank of England, also regularly use information from the LFS. Various other departments such as HM Revenue and Customs, Welsh Government and Scottish Government are also regular users of the LFS and APS.

Other users include local authorities, the CBI and other employer organisations, the TUC and individual trade unions, by labour market analysts in the City, economics correspondents in the broadsheet newspapers, and researchers in a wide variety of

other organisations ranging from the Unemployment Unit to the Institute for Employment Studies, from major retailers to solicitors

The Government Statistical Service is committed to providing statistical information as a basis for informing the wider public debate.

This article has described many, but not all, of the enormous range of uses to which Labour Force Survey results are used.

SECTION 17 - LFS DISSEMINATION AND PUBLICATIONS

17.1 DISSEMINATION

Availability of electronic data

The following LFS data exist electronically:

- annual LFS data for each survey from 1984-91
- annual "time series" database covering period 1984 to 1991
- each quarter's results since spring 1992 on both a regional and sub-regional basis
- household databases for Spring 1990, Spring quarters 1992-95 and Spring and Autumn quarters from 1995 onwards
- local area (counties, LADS, TECS, LECs) tabulations from spring 92 onwards (limited set of variables).
- 2 quarter and 5 quarter longitudinal datasets from winter 92/93. (available as SPSS files with a limited set of variables).

Services available from ONS

The LFS Data Advice and Relations Team (DART) can provide clients with tabulations from the LFS. Tables can be provided in Excel.

☎ LFS DART

socialsurveys@ons.gov.uk

Provision of these services may attract a charge, which will be detailed at the time of application.

The ONS Secure Research Service

The Secure Research Service (SRS) at the ONS gives accredited researcher secure access to special deidentified LFS/APS data to work on research projects for the public good.

For questions regarding project access levels, please contact srs.customer.support@ons.gov.uk.

The UK Data Archive

Formerly known as the ESRC Data Archive, based at the University of Essex, hold copies of all LFS databases. Academic users can access the data at specially agreed rates.

www.data-archive.ac.uk

Nomis®

The National On-line Manpower Information System holds local area LFS data.
　☎ Nomis®, 0191-334-2680
　✉ info@nomisweb.co.uk
　www.nomisweb.co.uk

17.2 PUBLICATIONS OF LFS RESULTS

The main Labour Market release (which is monthly) can be found here:
http://www.ons.gov.uk/ons/rel/lms/labour-market-statistics/index.html
This contains links to regularly published labour market releases including:
- Regional Labour Market statistics
- Public Sector Employment
- Young People who were Not in Employment, Education or Training (NEET)
- Working and Workless Households

As well as recently published reports on Labour Market topics.

Summaries and publications for the Labour Market can be found here:
*https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentand
employeetypes*

Labour market articles and reports on revisions and guides can be found here:
*https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentand
employeetypes/methodologies/labourmarketarticlesandreports*

Previous publications which used information from the LFS include:
- Social trends, (though this was discontinued in 2012)
  http://www.ons.gov.uk/ons/rel/social-trends-rd/social-trends/index.html
- Regional trends (discontinued in 2011)
  http://www.ons.gov.uk/ons/rel/regional-trends/regional-trends/index.html
- Labour market trends (which contains tables of historical LFS data, along with other labour market data) between 2001 and 2006
  http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/lms/labour-market-trends--discontinued-/index.html

17.3 ARTICLES ON LFS METHODOLOGY IN OTHER PUBLICATIONS

In addition, methodological issues and developments have been reported in a number of published articles[35]:

The ONS Methodology Working Series:

---

[35] The 2011 version of this user guide lists the articles from 1989 to 2005, this version of the user guide just list the most relevant ones.

https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries

-'Guide to calculating standard errors for ONS Social Surveys', M.Greenaway and B.Russ, No 9

The GSS Methodology Series:
http://www.ons.gov.uk/ons/guide-method/method-quality/specific/gss-methodology-series/index.html

- 'Methodological Issues in the production and analysis of longitudinal data from the Labour Force Survey', P.S. Clarke and P.F. Tate, *GSS Methodology Series*, No. 17, 1999

- 'Variance estimation for Labour Force estimates of Level and change', D.J Holmes and C.J Skinner, *GSS Methodology Series*, No 21, 2000

The Survey Methodology Bulletin:
http://www.ons.gov.uk/ons/guide-method/method-quality/survey-methodology-bulletin/index.html

- ' Interview mode effects in the UK local Labour Force Survey (LFS)', D. Elliot, L. Rainford and J. Eldridge, Survey Methodology Bulletin No. 57, March 2006

- 'Improving migrant participation in the Labour Force Survey (LFS): a review of existing practices in European Union member states', W. Barnes, Survey Methodology Bulletin No. 63, September 2008

- 'Improving migrant participation in the Labour Force Survey (LFS): non-response and attitudes of non-English-speaking migrants to participation', M. Thomas, Survey Methodology Bulletin No. 63, September 2008

- 'Non-response bias in the Labour Force Survey (LFS)', G. Bright, W. Barnes and D. Fletcher, Survey Methodology Bulletin No. 64, March 2009

- 'Pruning the labour force survey: removing respondents aged 75 and above from the Waves 2–5 interviews', K. Ashworth, M. Greenaway and P. Smith, Survey Methodology Bulletin No. 68, March 2011

- 'Design, implementation and testing of Labour Force Survey (LFS) questionnaire features in an online mode – results from the 2010/11 internet pilots', M. Portanti and L. Wilson, Survey Methodology Bulletin No. 70, March 2012

- 'Labour force refusal follow-up study', L. Wilson and N. Parry-Langdon, Survey Methodology Bulletin No. 71, September 2012

- 'A review of the Labour Force Survey longitudinal weighting methodology', F.Ariyibi and M. Greenaway, Survey Methodology Bulletin No. 74, Autumn 2015

17.4 ALTERNATIVE SOURCES OF LABOUR MARKET DATA

A guide to Labour Market Statistics can be found here:
http://www.ons.gov.uk/ons/rel/lms/labour-market-guidance/guide-to-labour-market-statistics/guide-to-lm-statistics.html
This contains information about the different sources used to measure many different aspects of work and jobs and to provide an insight into the economy.

Employment

The Workforce in Employment quarterly surveys of employers provides figures for employees in employment. They count jobs not people and are particularly valuable for providing estimates for specific industry sectors. A comparison of jobs from the two surveys - combining main and second employee jobs for the LFS - shows that the workforce series gives lower estimates of the number of employee jobs in the economy than the LFS, and that the gap has widened in recent years. For more details see 'Comparison of sources of employment data', *Labour Market Trends*, December 1997.

Unemployment

Figures for claimant count are produced every month from records of people claiming unemployment related benefits. These figures are available quickly and can be provided for small areas. The claimant count measure of unemployment and ILO unemployment are defined in different ways, and whilst there is some overlap between them, they could not be expected to correspond exactly. ILO unemployment tends to be higher than the claimant count with the smallest gap (72,000) occurring in spring 1993 when both measures were near a peak, and the largest (422,000) in spring 1990 when both measures were near a trough. For more details see 'LFS estimates of claimants of unemployment-related benefits: results of an ONS record linkage study'; Penny Pease, *Labour Market Trends*, November 1997.

Earnings & Hours

Statistics from the New Earnings Survey (NES) were published for the last time in 2003. A new survey, the Annual Survey of Hours and Earnings (ASHE) has been developed as a replacement. The ASHE survey is based on the same coding frame as NES. The new methodology includes improved coverage of employees and weighting of earnings estimates. The data and variables collected remains broadly the same.

Further information about these other surveys can be found here:
http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/labour-market/index.html

.

SECTION 18 - LFS DATA FOR SMALL SUB-GROUPS: ANNUAL DATABASES AND AVERAGING OVER SEVERAL QUARTERS

Introduction

The quarterly Labour Force Survey (LFS) contains information based on about 37,000 responding households in the United Kingdom and for many analyses this is sufficient. For small groups in the population such as ethnic minorities and young people, and for small area analyses in general, a single LFS quarter does not provide a large enough sample to give reliable estimates when cross-tabulations of several variables are required. The average of a larger sample contacted over a longer period will provide estimates of greater precision. Users may also wish to calculate annual average values for their own sake, or in order to smooth out seasonal or other variation.

Methods of combining LFS samples

A single LFS quarter includes information about some 100,000 individuals. This is sufficient to allow a wide range of labour market, educational and demographic analyses, but does not provide a large enough sample to give reliable results in detailed cross-tabulations for small population sub-groups or for local areas. The nature of sampling variability means that the smaller the group whose size is being estimated, the (proportionately) less precise that estimate is. Basing estimates on data for a larger sample can increase precision.

In order to meet the demand for more data at local authority district level, ONS released the first annual LFS Local Area Database (LADB) in May 1996 which covered March 1994 to February 1995. These databases contain a range of key variables together with a unitary authority local authority district (UA/LAD) identifier, which allows users to carry out cross-tabulations at local level for the first time. In 2000 to 2004 onwards additional boost surveys have also been added and conducted to increase the size of the LADB which were renamed the Annual Local Labour Force Survey datasets from 2000 to 2004 and the Annual Population datasets from 2004 onwards. The LFS user guide volume 6 provides some more information

Annual databases
To explain the concept of an annual database, it is first necessary to describe the panel design of the LFS, usually described as the 'wave' structure of the sample. Each quarter a new group (wave) of households is selected and its members interviewed for the first time. The same people are interviewed four more times at quarterly intervals. In any one quarter, one wave will be receiving their first interview, one wave their second and so on. Up to 1999/2000, the annual LADB was created by taking waves 1 and 5 from each of four consecutive quarters.

Selecting waves 1 and 5 ensures that each respondent is included once and only once - a database of 'distinct cases'. The selected records are weighted according to definitive mid-year population estimates. As a result of the larger sample size (60 per cent more than a quarterly LFS database), the standard errors of estimates (the usual measure of sample variation) from the local area database are about 80 per

cent of those for estimates for a single quarter. This resulted in the minimum publication threshold being reduced from 10,000 for a quarter, to 6,000 for annual data.

From March 2000, there was a boost to the sample in England - a partnership project between the ONS, the Department for Work and Pensions (DWP) and the Department for Education and Skills (DfES). More details including a summary publication, fact sheets for UA/LADs and counties are available on the National Statistics website - http://www.ons.gov.uk/ons/index.html. Respondents in the boost are interviewed once a year for four years. The increase in sample meant that for some UA/LADs, the publication threshold was reduced to 2,000 or 4,000.

From March 2001, there was a similar boost to the LFS sample for Wales – a partnership project between ONS and Welsh Assembly Government. Again, the increased sample resulted in lower publication thresholds for all the UAs in Wales, some as low as 1,000.

From March 2003 there was a similar boost to the LFS sample in Scotland – a partnership project between ONS and Scottish Executive.

From January 2004 there was an additional boost in England and the LADB (or ALALFS) was renamed the Annual Population Survey (APS). The boost was subsequently discontinued in January 2006 but the survey name remained as the APS. The APS is published quarterly. The overall sample size of the APS is approximately 146,000 individuals.

Annual averages
An alternative method is simply to add together the weighted estimates for four consecutive quarters (from the quarterly LFS databases or the time-series database) and divide by four. Such estimates will be averaging both the numerator and denominator. Therefore, the estimate from a 4 quarter average as an estimate of an annual average, will be inconsistent with the preferred APS estimate, as it will differ in certain key aspects. Firstly, the 4 quarter average is exactly that, the average of the 4 quarterly estimates. Therefore, if in say one quarter there were more people in the population than in another quarter, each quarter would be given the same weight in the calculation of the 4 quarter average. In the APS estimate, the annual estimate would give the quarters with the highest population more weight.

Due to the panel design of the LFS, databases for consecutive quarters have approximately 80 per cent of their samples in common. Hence the calculation of the precision (standard errors) of annual averages is not straightforward, it is not simply the sum of the four quarterly variances divide by 16. The improvement in accuracy or precision, (i.e. lower standard errors) arising from averaging estimates over a year is illustrated by the reduced thresholds for estimates regarded as sufficiently reliable for publication shown below.

<u>Minimum publication thresholds</u>

	Minimum publication level	95 per cent confidence interval
One quarter	10,000	+/- 4,000
Four quarters	6,000	+/- 2,640
Eight quarters	4,000	+/- 1,600
12 Quarters	3,000	+/- 1,200
16 Quarters	2,000	+/- 800

18 Reliability of averaged quarterly estimates

As a consequence of the overlap in samples between quarters, the precision of annual averages varies according to the correlation between responses from the same individuals in different quarters. Where responses are likely to change between quarters, as with economic activity variables, the gain from averaging is greater than with an annual database of distinct cases because the latter excludes additional information which has been collected about respondents in waves 2 to 4 each quarter. For variables which cannot change between quarters, such as ethnic origin, there is still a clear advantage, in terms of sampling error, in averaging data over a year compared with using a single quarter's data, although an annual database of distinct cases would produce somewhat more precise estimates than simple annual averages. In practice, users will `often be combining ethnic origin with labour market or education variables, so the difference between annual averages and annual database estimates would not be as great as this suggests.

Social Survey Directorate of ONS has provided some guidance on the level of estimates which would have the same relative precision as estimates of 10,000 in single quarter's LFS database. In other words, these are the smallest estimates which are considered reliable enough to use, based on a criterion of a maximum of 20 per cent coefficient of variation (the standard error as a percentage of the estimate). The recommended lower limits for reliable data for averages of successive quarters are shown in on the previous page.

The recommended thresholds are based on the quarter-on-quarter correlation exhibited by total employment and unemployment. They have been rounded up to the next 1,000, partly for simplicity and partly to allow for the additional loss of precision in variables which have virtually 100 per cent correlation. The calculation incorporates some design effects, to reflect the fact that the sample is clustered within addresses. However, these design effects vary widely for different ethnic groups and estimates up to two times these value may, for some groups, have confidence intervals as wide as those given on the previous page.

However, since the introduction of the boost surveys, increasing the annual datasets from 180,000 to 360,000 individuals, it is now recommended that the APS datasets are used always in preference from 4 quarterly average.

Other benefits and drawbacks of averaging

Annual averages can be calculated for any consecutive four-quarter period but combining numbers of quarters other than multiples of four is not recommended unless the data are seasonally adjusted. It is not advisable to calculate averages over periods where the questionnaire has changed.

For small area estimates there are other considerations than sampling errors when selecting a method of increasing the available sample size on which to base estimates. Carrying out the weighting of sample estimates to population totals as a last stage, as is done for the UA/LAD in the APS, is likely to produce more robust estimates for small areas than the averages of estimates from independently weighted-up quarterly databases. The UA/LADs also have the benefit of being weighted to the definitive mid-year population estimates rather than to projections as used for the quarterly databases. Hence, for the variables included, they may still be the better source for county estimates, although these are available from the quarterly databases and hence could be averaged.

Data for small sub-groups and areas always need to be interpreted with great caution. Differences between annual averages for different periods present a special problem since again there is some overlap in the samples for consecutive periods. In particular, drawing comparisons between annual averages for very small groups for periods ending one quarter apart (e.g. spring 2001 - winter 2001/02 compared with summer 2001- spring 2002) is not recommended.

For examining detailed characteristics of special groups, such as ethnic minorities, data could also be combined over two, three or more years. However, such estimates would be centred on a point a year or more in the past, making this less suitable for rapidly changing variables.
Whilst there are advantages to using annual averages, flexibility and timeliness, from 2000/01, the increased annual sample has resulted in more reliable estimates. As a result, ONS dramatically increased the amount of annual LFS data available free for local areas on Nomis®. Due to confidentiality constraints, ONS had to withdraw the publicly – available LADBs. However, annual LFS data are available, unrounded and unsuppressed with a disclaimer, from the Sub-National Data Service – contact LFS.dataservice@ons.gov.uk .
A charge may be made for this service. The APS datasets also became available from January 2004.

Non-standard variables

Earnings information is only available up to winter 1996/97 from fifth wave interviews, from spring 1997 it is available from first and fifth wave interviews. It is recommended that single-quarter estimates up to winter 1996/7 based on weighted-up estimates of less than 60,000 employees should not be used. From spring 1997 the threshold is 30,000 employees. When four quarters' earnings data are pooled, the threshold is unchanged as there is no overlap in the data. For annual databases, the thresholds were lower and from 1997/98, the thresholds for earnings data from the annual database are the same as for all other variables.

Some LFS variables are only available in two quarters each year. The comparative standard errors and reliability thresholds for multi-quarter averages quoted here do not apply to such variables. If there is demand, ONS would consider the appropriate guidance for such variables.

For variables which are only available once a year (for example, trade union membership), or indeed for other variables, users may wish to calculate averages of quarters a year apart. Here, the overlap (for the quarterly LFS conducted from 1992 onwards) is about 16 per cent. The appropriate thresholds for publication in this case are: two years: 6,000; three years: 4,000; four or five years: 3,000.

Calculation methods

Annual averages at national, regional and larger local authority level can be calculated by extracting tabulations of the same data for four quarters and placing them in a spreadsheet for manipulation. Care needs to be taken to check that the variables are consistently defined in all the quarters to be averaged.
Uses of multi-quarter averages

The value of averaging LFS estimates over four or more quarters is that the larger number of responses on which the estimate is based results in a more reliable estimate. This in turn means that smaller estimates become sufficiently reliable to use. One particular topic where
annual averages are already proving to be of great value for this reason is ethnic minority groups. By means of annual averaging, users are able to analyse smaller ethnic minorities and/or smaller sub-groups, such as age groups.

One statistic of great interest among policy makers and ethnic minority special interest groups is the youth unemployment rate for different ethnic groups. Annual averages can also make more regional estimates for ethnic minorities available. Aggregation over several years would allow analysis of the ethnic minority populations and other groups and regions in more depth than is possible from four quarters. Such estimates, while not fully up-to-date, would nevertheless be more recent than the 1991 population census. An example of information which is only available from the LFS is people who do unpaid work for a family business. The number of these is small, and up to now no analysis of the types of occupations they are employed in has been published by *ONS.*

Conclusion

One of the aims of ONS is to make better use of the statistics collected. Averaging LFS data over several quarters enables reliable estimates to be obtained for smaller groups than are available from the individual quarterly databases.
In addition, the increased annual samples have improved the reliability of LFS data for small areas. A wider range of data are now available free on the National Statistics and Nomis® websites and also from the Sub National Data Service. The timeliness of annual data has also been improved.

ANNEX A – PURPOSE LEAFLET

Office for
National Statistics

Labour Force Survey

The Office for National Statistics is the UK's largest independent producer of official statistics and the recognised national statistics institute of the UK.

We would like you to take part in a national study. The Labour Force Survey (LFS) is a survey of the employment circumstances of the UK population. This study is where the official employment and unemployment rates come from.

Below you can see what we found when we ran this study before. The UK may have changed since we ran the last study, so now we need your help to update this information.

You may not know...

JOBS   Between July and September 2020, redundancies increased to a record high of 314,000.

The employment rate for 16-64 year olds was 78% for men and 72% for women (August – October 2020).

Pre-lockdown anxiety was high. It was the highest seen since tracking began in 2012 (January – March 2020).

The full statistics and methodology behind this information can be found on our website:

🌐 www.ons.gov.uk/householdstudies

About the study

This study is where employment and unemployment rates come from. The data affects the interest rates. This impacts mortgage debts, pensions, and savings. The study also covers retirement, higher education and looking after the family or home. Recently the study has informed labour market policies for people who have been on furlough during the pandemic.

Whoever you are and whatever you do, we are keen to hear from you.

How to find out more

Online

Our website contains more information about how you were chosen and what we do with your data:

🌐 www.ons.gov.uk/householdstudies

You can also get in contact through our social media pages.

f www.facebook.com/ONS

🐦 www.twitter.com/ONS

By telephone

You can call us free on 0800 298 5313
Text Relay Service (18001) 0800 298 5313

Our phone lines are open:

Monday to Friday: 8am to 7pm
Saturday: 8am to 1pm

LFS_LT_ENG 10/21

ANNEX B – DERIVATION OF STANDARD ERRORS ON THE LFS

An approximation for the standard error of an estimated count (for example, the number of people of a particular characteristic)

We start from the formula for the standard error (SE) of an estimated proportion, p, which is given by:

$$SE(p) = \sqrt{\frac{p(1-p)}{n}}$$ where n is the sample size.

If the survey design is complex, an additional factor should be included in the formula, the Design Factor, DEFT. DEFT, defined as the square root of the Design Effect (DEFF), is the ratio of the standard error under the complex design to that under a simple random sample design.

Thus, we get: $SE(p) = DEFT \times \sqrt{\frac{p(1-p)}{n}}$

If N is the total size of the population (i.e. including people of all characteristics), and M is the estimated count of people with a particular characteristic, then we estimate $p = \frac{M}{N}$, or equivalently, $M = N \times p$

It follows that

$$SE(M) = N \times SE(p)$$
$$= DEFT \times N \times \sqrt{\frac{p(1-p)}{n}}$$
$$= DEFT \times N \times \sqrt{\frac{M}{N}\left(\frac{N-M}{N}\right)\frac{1}{n}}$$
$$= DEFT \times \sqrt{\frac{M(N-M)}{n}}$$

An approximation follows, which is appropriate for estimates of relatively small populations. If we now assume:
- that M is relatively small in comparison to N (i.e. that $(1 - p)$ is close to 1),
- the design factor is also close to 1,
- and by replacing the design weight, N / n, by its current value (approximately equal to 580)

then we get

$$SE(M) \approx \sqrt{M\frac{N}{n}} = \sqrt{M \times 580}$$

Equivalently,
$SE(M_T) \approx \sqrt{M_T \times 0.58}$, where $M_T = M/1000$ is the estimated count in thousands.

Therefore, an approximate 95% confidence interval for M_T is given by

$$M_T \pm 2\sqrt{M_T \times 0.58}$$

i.e.

$$M_T \pm \sqrt{M_T \times 2.32}$$

Standard error of a ratio

The sampling variance in the survey was estimated by treating each household as a primary sampling unit (PSU). The method used to estimate the standard error of a ratio R=Y/X (where X might, for example, be the number of men aged 16 and over who are economically active and Y might be the number of them who are unemployed) was that of successive differences. The formulae used were:

$$Variance\ (Y) = \sum_s \frac{N_S}{2(N_S-1)} \sum_{p=2}^{N_S} (Y_{s,p} - Y_{s,p-1})^2$$

$$Variance\ (X) = \sum_s \frac{N_S}{2(N_S-1)} \sum_{p=2}^{N_S} (X_{s,p} - X_{s,p-1})^2$$

$$Covariance\ (X,Y) = \sum_s \frac{N_S}{2(N_S-1)} \sum_{p=2}^{N_S} (X_{s,p} - X_{s,p-1})(Y_{s,p} - Y_{s,p-1})$$

$$StandardError(R) = \frac{1}{X}\sqrt{(VAR(Y) - 2RCOV(XY) + R^2VAR(X))}$$

Where s represents the stratum, p the PSU, I the individual case, $X_{s,p}$ and $Y_{s,p}$ the PSU totals, and N_s the number of PSUs in stratum s.

ANNEX C – LABOUR FORCE SURVEY STANDARD ERRORS: OCTOBER TO DECEMBER 2023, UNITED KINGDOM

Economic activity status

Characteristic	Estimate (%)	Standard Error	Design Factor	Unweighted Base
All people aged 16+ by Economic Activity:				
Economically Active	62.82	0.287	1.1144	35,210
All in Employment	60.41	0.2941	1.1285	35,210
Employees	52.25	0.3078	1.1563	35,210
Self-Employed	37.18	0.287	1.1144	35,210
ILO Unemployed	7.93	0.1811	1.2575	35,210
Economically Inactive	2.41	0.1121	1.3731	35,210
Men aged 16+ by Economic Activity:				
Economically Active	66.92	0.3779	1.0377	16,687
All in Employment	64.18	0.394	1.0615	16,687
Employees	53.59	0.4232	1.0963	16,687
Self-Employed	33.08	0.3779	1.0377	16,687
ILO Unemployed	10.27	0.2781	1.1834	16,687
Economically Inactive	2.74	0.1679	1.3282	16,687
Women aged 16+ by Economic Activity:				
Economically Active	58.95	0.377	1.0431	18,523
All in Employment	56.86	0.3869	1.0632	18,523
Employees	50.99	0.403	1.0972	18,523
Self-Employed	41.05	0.377	1.0431	18,523
ILO Unemployed	5.73	0.213	1.247	18,523
Economically Inactive	2.09	0.147	1.3996	18,523

People aged 16+ in employment by type of employment

Characteristic	Estimate (%)	Standard Error	Design Factor	Unweighted Base
All people aged 16+ in employment by type of employment:				
Full-time	65.8	0.3987	1.2257	20,915
Part-time	0.12	0.0272	1.1229	20,915
Government employment and training	20.6	0.3262	1.1762	20,915
Unpaid family worker	0.25	0.0443	1.2971	20,915
Men aged 16+ in employment by type of employment:				
Full-time	73.3	0.508	1.1885	10,400
Part-time	0.17	0.046	1.1509	10,400
Government employment and training	10.11	0.3451	1.1844	10,400
Unpaid family worker	0.34	0.0801	1.4318	10,400
Women aged 16+ in employment by type of employment:				
Full-time	57.81	0.5949	1.2361	10,515
Part-time	0.08	0.028	1.0495	10,515
Government employment and training	31.77	0.5436	1.1983	10,515
Unpaid family worker	0.16	0.0325	0.844	10,515

People aged 16+ in employment by industry sectors

Characteristic	Estimate (%)	Standard Error	Design Factor	Unweighted Base
All people aged 16+ in employment by industry sector:				
Agriculture, forestry and fishing	0.93	0.0837	1.2748	20,915
Energy and water	1.67	0.1101	1.254	20,915
Manufacturing	8.1	0.2462	1.3157	20,915
Construction	6.38	0.2111	1.2595	20,915
Distribution, hotels and restaurants	15.68	0.3348	1.3429	20,915
Transport and communication	9.63	0.2665	1.3174	20,915
Banking and finance	19.13	0.3565	1.3219	20,915
Public admin, education and health	32.53	0.4028	1.2538	20,915
Other services	5.94	0.2103	1.2975	20,915
Men aged 16+ in employment by industry sector:				
Agriculture, forestry and fishing	1.37	0.1364	1.2144	10,400
Energy and water	2.51	0.1815	1.2017	10,400
Manufacturing	11.57	0.4033	1.3048	10,400
Construction	10.64	0.3755	1.2602	10,400
Distribution, hotels and restaurants	15.29	0.4477	1.2874	10,400
Transport and communication	13.76	0.4164	1.2511	10,400
Banking and finance	19.44	0.4744	1.2407	10,400
Public admin, education and health	20.1	0.4845	1.2512	10,400
Other services	5.33	0.283	1.304	10,400
	1.37	0.1364	1.2144	10,400
Women aged 16+ in employment by industry sector:				
Agriculture, forestry and fishing	0.45	0.0666	1.0172	10,515
Energy and water	0.77	0.1037	1.2137	10,515
Manufacturing	4.41	0.2445	1.2216	10,515
Construction	1.85	0.16	1.2199	10,515
Distribution, hotels and restaurants	16.1	0.457	1.2761	10,515
Transport and communication	5.24	0.2904	1.3372	10,515
Banking and finance	18.81	0.4946	1.2987	10,515
Public admin, education and health	45.77	0.5966	1.2289	10,515
Other services	6.59	0.3045	1.2593	10,515

Economically active people 16+ by sex who were ILO unemployed

Characteristic	Estimate (%)	Standard Error	Design Factor	Unweighted Base
Economically active people aged 16+ who were ILO unemployed:				
All persons 16+	3.83	0.177	1.372	21,548
Men 16+	4.1	0.249	1.3278	10,758
Women 16+	3.54	0.2476	1.3996	10,790
Economically active people aged 16-17 who were ILO unemployed:				
All persons 16-17	20.09	2.5685	1.1624	333
Men 16-17	21.25	3.5273	1.0675	159
Women 16-17	19.09	3.6991	1.2453	174
	20.09	2.5685	1.1624	333
Economically active people aged 18-24 who were ILO unemployed:				
All persons 18-24	10.40	0.9272	1.2004	1,574
Men 18-24	13.46	1.4141	1.1909	826
Women 18-24	7.02	1.1083	1.1766	748
	10.40	0.9272	1.2004	1,574
Economically active people aged 25-34 who were ILO unemployed:	13.46	1.4141	1.1909	826
All persons 25-34	3.78	0.4027	1.2763	3,697
Men 25-34	3.81	0.5524	1.2092	1,774
Women 25-34	3.76	0.5653	1.2930	1,923
	3.78	0.4027	1.2763	3,697
Economically active people aged 35-49 who were ILO unemployed:				
All persons 35-49	2.71	0.2674	1.3794	7,077
Men 35-49	2.36	0.3318	1.2779	3,461
Women 35-49	3.06	0.4270	1.4817	3,616
Economically active people aged 50-64 who were ILO unemployed:				
All persons 50-64	2.16	0.2101	1.2542	7,398
Men 50-64	2.35	0.3055	1.2290	3,662
Women 50-64	1.96	0.2757	1.2261	3,736
	2.16	0.2101	1.2542	7,398
Economically active people aged 65+ who were ILO unemployed:				
All persons 65+	1.28	0.4631	1.0641	876
Men 65+	1.32	0.7298	1.2665	593
Women 65+	1.30	0.4210	1.2057	1,469

People aged 16+ in employment by region of usual residence (GOR)

Characteristic	Estimate (%)	Standard Error	Design Factor	Unweighted Base
All persons aged 16+ in employment:				
North East	55.13	1.4215	1.0734	1,523
North West	59.32	0.8487	0.9537	3,484
Merseyside	55.50	1.4641	1.1960	1,548
Yorkshire & Humberside	58.53	1.0547	1.1539	2,502
East Midlands	57.78	0.8793	0.5536	4,535
West Midlands	60.42	1.1091	1.2677	2,507
Eastern	52.92	2.6890	1.5016	471
London	59.12	0.9782	1.0681	2,809
South East	58.53	0.9417	0.9705	2,811
South West	59.42	0.9896	1.1227	2,855
Wales	60.32	0.8307	0.9782	3,122
Scotland	67.66	1.0356	1.5002	2,438
Northern Ireland	62.77	0.7538	1.0865	4,605
All persons aged 16+ ILO unemployed:				
North East	2.33	0.5088	1.2673	1,523
North West	1.65	0.2558	1.1069	3,484
Merseyside	2.00	0.5185	1.5052	1,548
Yorkshire & Humberside	2.75	0.4374	1.4417	2,502
East Midlands	1.52	0.2396	0.6097	4,535
West Midlands	2.37	0.3662	1.3457	2,507
Eastern	3.91	1.1753	1.6911	471
London	1.98	0.3245	1.2490	2,809
South East	3.18	0.4181	1.2105	2,811
South West	2.45	0.3868	1.3935	2,855
Wales	2.10	0.2968	1.1932	3,122
Scotland	2.74	0.3713	1.5426	2,438
Northern Ireland	2.50	0.3269	1.4593	4,605
All persons aged 16+ economically inactive:				
North East	42.54	1.3729	1.0430	1,523
North West	39.03	0.8429	0.9538	3,484
Merseyside	42.51	1.4386	1.1814	1,548
Yorkshire & Humberside	38.72	0.9946	1.1005	2,502
East Midlands	40.70	0.8525	0.5396	4,535
West Midlands	37.21	1.0972	1.2687	2,507
Eastern	43.17	2.6861	1.5116	471
London	38.90	0.9751	1.0736	2,809
South East	38.30	0.9250	0.9661	2,811
South West	38.13	0.9736	1.1166	2,855
Wales	37.58	0.8068	0.9597	3,122
Scotland	29.61	1.0130	1.5038	2,438

Northern Ireland	34.73	0.7167	1.0489	4,605

People aged 16+ by ethnic origin

Characteristic	Estimate (%)	Standard Error	Design Factor	Unweighted Base
White				
All persons 16+	85.49	0.3207	1.7079	35,166
Men 16+	85.88	0.3974	1.4733	16,665
Women 16 +	85.13	0.3616	1.3823	18,501
	85.49	0.3207	1.7079	35,166
Mixed				
All persons 16+	1.48	0.1060	1.6486	35,166
Men 16+	1.39	0.1498	1.6526	16,665
Women 16 +	1.56	0.1378	1.5134	18,501
Asian or Asian British				
All persons 16+	6.93	0.2542	1.8772	35,166
Men 16+	6.97	0.3105	1.5738	16,665
Women 16 +	6.89	0.2770	1.4880	18,501
Black or Black British				
All persons 16+	3.37	0.1817	1.8891	35,166
Men 16+	3.07	0.2151	1.6113	16,665
Women 16 +	3.65	0.2081	1.5084	18,501
Chinese				
All persons 16+	0.72	0.0791	1.7586	35,166
Men 16+	0.65	0.0988	1.5872	16,665
Women 16 +	0.78	0.1023	1.5814	18,501
Other				
All persons 16+	2.02	0.1381	1.8425	35,166
Men 16+	2.04	0.1678	1.5319	16,665
Women 16 +	1.99	0.1575	1.5340	18,501