

Transformed Labour Force Survey background user guide

14 November 2022

Contents

Section 1: Background on the transformation of the Labour Force Survey	2
Purpose	2
History	2
Context of the transformation	2
Development of the transformed survey	3
Section 2: Survey design	4
Target population	4
Sampling frame and selection	4
Wave structure	4
TLFS Core and TLFS Plus	4
Longitudinal design	5
Rolling reference week	6
Mode	7
Adaptive survey design	7
Responsive survey design	8
Item non-response rates	8
Section 3: Weighting	9
Section 4: Imputation	10
Section 5: Quality	11
Section 6: Dissemination	12
Government departments	12
Secure Research Service (SRS)	12
UK Data Service (UKDS)	12

Section 1: Background on the transformation of the Labour Force Survey

Purpose

The current 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 Annual Population Survey (APS) is built using LFS data to create larger datasets to allow more granular analysis at the expense of timeliness. It is typically used for both labour market and demographic analysis.

History

The first LFS in the UK was conducted in 1973. Initially the survey was carried out every two years. However, over time, as its use by UK government departments increased, the regularity of data capture increased. An annual survey was conducted from 1984, and a quarterly survey was carried out from 1992 in Great Britain and 1994 in Northern Ireland.

Alongside these changes in regularity, the panel design was also developed in the early 1990s. This meant that respondents were then asked some of the same questions for five consecutive quarterly surveys. In addition, new topics such as earnings were introduced.

Extra respondents were included as an annual enhancement to the sample size of the LFS in 2000. Those local boost samples were introduced to improve labour market information at local level. This was further improved in 2004 to create the Annual Population Survey (APS), which brings together all responses to the quarterly LFS collected over the course of a year to create a larger analytical dataset.

Both the current LFS and APS will be replaced by the transformed survey.

The full history and developments of the LFS are available in [Volume 1 – LFS background and methodology user guide \(PDF, 2.3MB\)](#). The current [Labour Force Survey user guidance](#) is also available.

Context of the transformation

In 2014, we published a [National Statistics Quality Review of the Labour Force Survey \(PDF, 1.21MB\)](#) containing some important recommendations to improve the design of the LFS and the quality of its outputs, for example, by:

- introducing online data collection
- exploring the potential of identifying priority areas in the field and allocating resources more flexibly to these areas
- moving to AddressBase as the sampling frame for residential households
- exploring the option of using a rolling reference week rather than a fixed reference week

The LFS transformation process has focused primarily on addressing these recommendations.

In addition, when working to introduce online data collection, research showed that some respondents found the questions difficult to understand. Historically, highly trained and

dedicated interviewers have played an essential part in helping respondents apply the questions to their specific circumstances. To account for this, the transformation has also focused on a respondent-centred design of questions.

Development of the transformed survey

The transformation of the Labour Force Survey (LFS) has primarily been conducted using a prototype survey instrument known as the Labour Market Survey (LMS).

Initial work included the adoption of a [respondent-centred design approach](#) for the development of survey questions for the LMS prototype.

In parallel, we have conducted a series of large-scale quantitative tests. These were mostly online-only tests of households, with a sample from across Great Britain. They focused on survey design factors such as respondent materials, incentives and contact strategies. Reports and findings on these tests are available online:

- [Labour Market Survey Response rate experiments Report for Test 1: Materials experiment \(PDF, 2.18MB\)](#)
- [Labour Market Survey Response rate experiments Report for Test 2, Tranche 1: Incentives experiment \(PDF, 2.32MB\)](#)
- [Labour Market Survey: technical report](#)
- [Labour Market Survey: comparative estimates report](#)
- [Labour Market Survey: characteristics report](#)

In March 2020, we launched the Transformed Labour Force Survey (TLFS) in response to the coronavirus (COVID-19) pandemic. At that time, the survey only allowed online response.

In 2021, we conducted a statistical and operational trial to determine the impact of two new approaches to survey response:

- **Knock-to-nudge:** This approach involves field interviewers knocking on a respondents' door to encourage response if they have not previously answered the survey.
- **Telematching:** This approach involves telephone interviewers calling respondents to encourage response and conduct interviews where the telephone number could be obtained via a telematching service provided by a third-party organisation.

Both new approaches had a positive impact on overall response rates. Additionally, they improved the representativeness of the achieved sample by including proportionately more working age, ethnically diverse and non-UK born respondents.

From February 2022, the ability for respondents to answer the survey by telephone was provided as an option across the UK. In November 2022, knock-to-nudge processes were rolled-out across the whole TLFS in Great Britain, with changes to the content of the questionnaire being implemented on an ongoing basis.

Section 2: Survey design

Target population

The Transformed Labour Force Survey (TLFS) targets UK residents in private households. The survey does not currently include communal establishments. One exception to this are students living in halls of residence who are included at their parental address.

Sampling frame and selection

Addresses are selected using systematic random sampling.

In England, Wales and Scotland, the sample for addresses is drawn from [AddressBase Premium](#). This address database owned by Ordnance Survey is comprised of local authority data, Royal Mail data and Ordnance Survey data, and is available to the Office for National Statistics under the [Public Sector Mapping Agreement](#).

In Northern Ireland, the sample for addresses is drawn at random from the Northern Ireland Statistics and Research Agency Address Register, which is a comprehensive list of addresses primarily based on the Land & Property Services [Pointer](#) database.

The sample is split and issued across a 13-week data collection period following calendar quarters. Each weekly sample, referred to as a “cohort”, is geographically representative.

Wave structure

The TLFS is a longitudinal survey consisting of five waves. It uses a rotational sampling design, whereby an address, once initially selected, is retained in the sample for a total of five consecutive quarters. We define Wave 1 as the first quarter an address is selected, Wave 2 as the second quarter an address is selected, and so on. Wave 5 is the last time that an address will be invited to take part in the TLFS.

Therefore, in any given quarter, interviews will be carried out across all five waves. For example, in the January to March 2022 quarter:

- a new cohort of addresses was interviewed for Wave 1
- the cohort of addresses sampled in October to December 2021 were invited for their Wave 2 interview
- the cohort of addresses sampled in July to September 2021 were invited for their Wave 3 interview
- the cohort of addresses sampled in April to June 2021 were invited for their Wave 4 interview
- the cohort of addresses sampled in January to March 2021 were invited for their Wave 5 interview

TLFS Core and TLFS Plus

The Wave 1 questionnaire is referred to as the “Master Wave”. It is currently designed to collect socio-demographic data and core labour market data.

All households invited to take part in the TLFS will be asked the TLFS Core questions focusing on core labour market content, while half of all households invited will be asked to complete an extended version of the survey which includes TLFS Plus questions.

This design aims to minimise respondent burden and improve data quality by keeping the core questionnaire as short as possible while recognising the data need for the TLFS Plus content. The sample size of the TLFS Core maximises the ability to analyse those variables at a granular geographic and characteristic level.

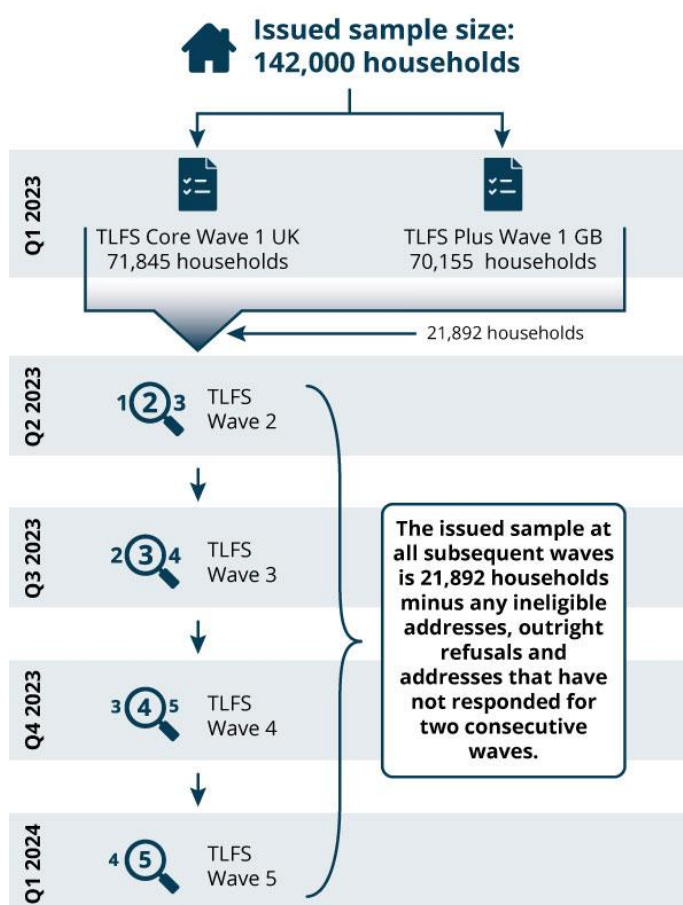
The only exception to this is for Northern Ireland sampled addresses which are currently all issued to TLFS Core only.

Longitudinal design

Of the 142,000 Master Wave addresses, 21,892 are rotated to Wave 2. These are randomly pre-selected from both TLFS Core and TLFS Plus addresses before Wave 1 data collection begins. A proportional number of cases are selected per country and region and, as they are pre-selected, they contain a mixture of both responding and non-responding addresses. Only outright refusals and ineligible are not rotated to Wave 2. “Outright refusals” refer to respondents who are not physically or mentally capable or refuse to respond to the survey and the interviewer feels that there is no chance of an interview at the current or in any future wave. “Ineligibles” refer to addresses that are not eligible to take part in the survey, such as commercial addresses, second homes, communal establishments, or vacant properties.

The wave structure and the longitudinal design are represented in Figure 1.

Figure 1: TLFS wave structure and longitudinal design



The exception to this is that all sampled addresses in Northern Ireland are rotated to Wave 2 as opposed to only a proportion of addresses in England, Wales and Scotland.

All addresses at Wave 2 onwards are issued the TLFS Core questionnaire, with the same questions asked at every wave. In future, we hope that some of the questions will only be asked where there has been a change in circumstance since the last wave to reduce respondent burden.

An estimation of the issued and achieved sample sizes for the TLFS Core and TLFS Plus designs is shown in Table 1 and Table 2.

**Table 1: Issued and estimated achieved samples for TLFS Core
Number of addresses per quarter**

	Issued sample	Estimated responses
Wave 1	142,000	52,398
Wave 2	21,892	4,957
Wave 3	7,699	3,233
Wave 4	5,112	2,408
Wave 5	3,310	2,062

Notes:

1. Wave 1 figures are calculated using average response rates since Quarter 2 (April to June) 2022. Response rates are expected to increase with the introduction of the knock-to-nudge approach.
2. 70,155 cases out of the 142,000 sample from the TLFS Core Wave 1 are issued the TLFS Plus.
3. 24,273 cases out of the 52,398 estimated responses from the TLFS Core Wave 1 are achieved through the TLFS Plus.

**Table 2: Issued and estimated achieved samples for TLFS Plus
Number of addresses per quarter**

	Issued sample	Estimated responses
Wave 1	70,155	24,273

Note:

1. The 70,155 cases are part of the overall 142,000 Wave 1 sample, not an additional sample.

Rolling reference week

The TLFS uses a “rolling reference week” for Wave 1, which is intended to reduce recall bias and improve respondent recall for key measures. For each sampled address, the rolling reference week is set to the week prior to the date on which the household started the survey. Once the reference week for a household has been defined, it remains static; if a household returns to the survey at a later date to enter further data, the reference week remains unchanged.

As the TLFS has a four-week collection period at Wave 1, fixing the reference period to a specific week may have introduced recall bias or increased respondent burden. For example, a household completing in week 4 that has to recall the number of hours they worked five weeks ago may need to consult documentation or provide an estimate rather than provide an accurate number. By making the reference week the week immediately prior to the interview, it reduces the burden and reduces the effort to recall accurate data.

From Wave 2 onwards, the TLFS uses a fixed reference week to maintain a fixed 13-week gap between waves and because the collection period is only two weeks compared to the four-week collection period at Wave 1.

Mode

The TLFS is an online-first, multi-mode survey.

Sampled addresses are sent a letter inviting them to take part in the survey. A 12-digit unique access code is provided to enable respondents to take part online. An option to complete via telephone is also available. The questionnaire is available in English or Welsh. A telephone interpretation service is available for those who may not be able to take part in those languages.

Sampled addresses that have not yet completed the survey are sent a reminder letter after 10 days and may also be encouraged to take part via telephone. All sampled addresses are telematched by a third-party company that provides mobile or landline telephone numbers for each address where available. Currently, telephone numbers are provided for around 30% of the sample. Our telephone interviewers then make outbound calls to these addresses asking respondents to take part in a telephone interview. Addresses sampled in Northern Ireland are not included in the telematching exercise.

For those who have not responded after two weeks, trained field interviewers may be used to conduct knock-to-nudge visits to sampled addresses to encourage online or telephone response through face-to-face engagement on the doorstep. On these visits, interviewers will also capture telephone numbers where possible to enable further follow-up via telephone if needed, for this wave or subsequent waves. This form of face-to-face engagement will be introduced in November 2022. The success of knock-to-nudge visits will be assessed to determine whether traditional face-to-face interviewing will be required in the future. Addresses sampled in Northern Ireland are not currently included in the knock-to-nudge visits but may be in future.

Adaptive survey design

Not all non-responding addresses will receive knock-to-nudge visits. Field interviewer resource will be targeted at addresses in areas in Great Britain with the lowest likelihood of response. This is part of our adaptive survey design, which uses information on how people with different demographics (or areas characterised by people with particular demographics) respond differently and adapts the survey design to reduce non-response error in the data produced.

A response propensity model has been developed using historical TLFS data to determine the characteristics of areas at Lower Super Output Area (LSOA) level least likely to

respond without intervention and to identify where knock-to-nudge visits are likely to be most effective in reducing non-response error.

A first iteration of the model found three variables and sub-categories to be strongly associated with non-response. Knock-to-nudge visits will therefore be targeted at areas with these characteristics: [Rural/urban classifications](#) (urban), [Age](#) (average age across LSOA is over 45 years old) and Index of Multiple Deprivation (most deprived LSOAs, deciles 1 to 4). The devolved administrations of Great Britain publish their own Index of Multiple Deprivation:

- [English indices of deprivation](#)
- [Welsh Index of Multiple Deprivation](#)
- [Scottish Index of Multiple Deprivation](#)

Responsive survey design

By implementing a responsive survey design, we can adapt the TLFS collection operation to meet changing needs, drive operational efficiencies and improve data quality. Using detailed timely management information on how the collection operation is performing, interventions can be put in place rapidly to address underperformance in particular areas or population groups or to respond to operational events.

This design will mature over time but is likely to include:

- the ability to increase or decrease interviewer capacity in particular areas
- engagement with community leaders to encourage response
- changes to letters or posting dates for example, to manage the impact of industrial action

Item non-response rates

As part of our quality assurance process, we will monitor non-response for individual variables to ensure that levels of missing data for key variables are within tolerance limits.

Many of the core labour market questions cannot be skipped in the questionnaire therefore item non-response rates will be low for these measures. If non-response for non-core variables exceeds tolerance levels, then we will review the design of these questions and attempt to identify the reasons for non-response.

If required, we will conduct additional research to improve the questions with the aim of increasing the response rate to those particular items. Information on imputation, which is a topic linked with item non-response, is available in [Section 4: Imputation](#).

Section 3: Weighting

The weights applied in the Transformed Labour Force Survey (TLFS) reflect the methodology used in the Labour Force Survey (LFS). This means users can refer to the current [Volume 1 – LFS background and methodology user guide \(PDF, 2.3MB\)](#).

There are small differences in some of the calibration groups. The LFS uses logistic regression to deal with non-response whereas the TLFS uses a pre-calibration step. The TLFS also calibrates to the LFS population totals.

Work is underway to test different weighting systems as the TLFS develops. We are aiming to compare estimation results with census results, current LFS results and other relevant administrative sources to ensure that we provide the most optimal estimation.

We are also working on developing a model-based approach to provide more timely estimates from the TLFS. At present, our Labour Market team only produces rolling-quarter national estimates as official statistics. We are planning on testing state space and Bayesian hierarchical models to assess if we can produce more breakdowns from the TLFS data (by sex, region, age groups, etc.) and whether we can produce estimates on a monthly basis rather than rolling quarterly.

We will provide further details on this approach in an updated version of this user guide once we have quality-assured the testing results and have a finalised proposal.

Section 4: Imputation

The Labour Force Survey uses a combination of last observation carried forward (LOCF) and nearest neighbour imputation method across the available person and household level files. The LOCF method uses the last observed value of an individual to impute that value if the next interview is a non-contact. A limitation of this method is that it assumes that the last observed value will remain valid at the time of the next (non-responding) interview. LOCF does not acknowledge variance (measurement errors or random fluctuations over time). The method can often inadvertently produce an overly stable picture of the actual situation and might underestimate flows.

The Transformed Labour Force Survey (TLFS) employs a missForest non-parametric imputation method for mixed type data. missForest is applicable to any type of input data and makes as few assumptions as possible about structural aspects of the data. In contrast to the nearest neighbour method which performs best for continuous data only and makes normal distribution assumptions, missForest is more suitable for the nature of the TLFS datasets. missForest uses machine learning random forest (RF) algorithms in an iterative fashion and as a non-parametric method it allows for interactive and non-linear as well as linear effects. missForest algorithms work in an iterative fashion by training an RF on observed values in the first step, followed by predicting the missing values and then proceeding iteratively until all columns with missing data are filled using RF. missForest also has stopping criteria to ensure that the best predicted values are selected. This means that as soon as the difference between the newly imputed data matrix and the previous one increases for the first time, the algorithm stops.

RF has been shown to perform well under barren conditions involving complex interactions, non-linear data structures and high dimensionality within datasets. This ensures our imputation approach is optimal to address possible data complexity in TLFS. RF rarely needs tuning parameters and hence is easy to use and needs no prior knowledge about the data because it is tested on the data themselves. This contrasts with the nearest neighbour method which is very sensitive to outliers and noise in the data. Academic work has shown that missForest outperforms other population imputation techniques including nearest neighbour methods. We tested 16 different imputation algorithms (including nearest neighbour) on random TLFS quarters and on all possible patterns of missingness and missForest produced the least amount of error.

We expect that the final system will involve LOCF imputation for demographic data that are not subject to change, and missForest imputation for other data types.

We will provide further details on the final approach in an updated version of this user guide once we have tested linking observations across waves.

Section 5: Quality

A larger sample size with more representative data is fundamental to delivering outputs with a higher degree of precision and producing estimates at a more granular geographic level.

To increase the quality of the data, the collection operation will focus on three areas listed in priority order:

1. Reduce bias: by reducing variability in response across geographic areas and across other area classifications such as Index of Multiple Deprivation and Output Area Classifications.
2. Reduce attrition: by retaining respondents through to Wave 5 and reducing the level of bias across each wave.
3. Improve response: by increasing overall response rates and achieving a minimum level of response in each local authority and region.

In addition, we will compare the proportional make-up of respondent characteristics (age, sex, disability, country of birth, tenure, ethnicity, occupation, industry) with best available estimates of the overall population to identify particularly under-represented groups and target interventions towards those groups where possible.

By manipulating the design of the collection operation to drive up the quality of the data collected through our adaptive and responsive survey design, the Transformed Labour Force Survey will better enable the production of higher quality, timely, more granular estimates that meet user needs.

Section 6: Dissemination

We will use a range of Transformed Labour Survey (TLFS) datasets across our analytical outputs, based on the analysis need. We will make any microdata that have been used in our published analyses available to users. We will do so using a range of mediums depending on the type of users.

We are currently developing the production of the following TLFS microdata, which will be replacing all current LFS and APS datasets:

- Rolling-quarter (including calendar-quarter) cross-sectional person basis
- Calendar-quarter cross-sectional household basis
- Two-quarter (2Q) longitudinal person basis
- Two-quarter (2Q) longitudinal household basis
- Five-quarter (5Q) longitudinal person basis
- Annual cross-sectional person basis
- Annual cross-sectional household basis

We will assess the need for pooled datasets covering wider time spans as the volume of TLFS data increases.

Government departments

Government departments will continue to receive identifiable Labour Force Survey (LFS) microdata through the transition every calendar quarter with their access and microdata specifications governed through our Data Access team. Microdata will be made available in .CSV format (including a data file and a metadata file). This can then be converted into a format of the user's choice such as SPSS, SAS, STATA, R and Python. Datasets are delivered to government departments to an agreed set of contacts via the secure MOVEit file transfer system, on each corresponding Labour Market release day.

Secure Research Service (SRS)

The SRS receive identifiable LFS microdata quarterly containing largely the same information and versions that are delivered to government departments. [More information about the SRS](#) is available.

UK Data Service (UKDS)

The [UKDS](#) receive non-identifiable "End User Licence" LFS microdata on a quarterly basis. These microdata are recoded for certain identifiable variables to a more aggregated level (for example, marital status, nationality, country of birth), and their specification contains a much-reduced list of variables compared to the SRS version. Microdata are delivered via the secure MOVEit file transfer system on each corresponding Labour Market release day.

We aim to revise the user guidance in spring 2023 including more details on what the different datasets cover. We will also provide further information on the methodological design and content of the survey as it develops.