

Investigating non-response to the Labour Force Survey using administrative data

Using Pay As You Earn data to identify and adjust for potential employment-related non-response to the Labour Force Survey. These are not official statistics.

Contact:
National Statistician's Analysis
Unit
NSO_Analysis_Unit@ons.gov.uk
+44 1329 444110

Release date:
28 September 2026

Next release:
To be announced

Table of contents

1. [Overview](#)
2. [The Labour Force Survey](#)
3. [Methodology for incorporating administrative data into survey estimation](#)
4. [Results of applying the methodology](#)
5. [Further work](#)
6. [Definitions](#)
7. [Data on investigating non-response to the Labour Force Survey using administrative data](#)
8. [Related links](#)
9. [Cite this page](#)

1 . Overview

The Labour Force Survey (LFS) is used to derive essential labour market indicators such as rates of employment, unemployment and economic inactivity.

As with many household surveys nationally and internationally, the LFS has experienced a decline in response rates over several decades, increasing the potential for non-response bias. While the current survey weights are designed to reduce non-response bias according to several characteristics that may be correlated with employment status, such as age, sex and geography, they may be further improved by incorporating administrative data specifically related to employment. Most people in employment are employees, so one such data source is Pay As You Earn (PAYE) Real Time Information (RTI).

This article reports experimental analysis showing how linked PAYE data may provide insights into patterns of employment-related non-response to the LFS. It also extends the current LFS survey weights to include PAYE information and assesses the impact on headline labour market estimates from 2019 to 2025 for Great Britain (data linkage is not yet fully in place for LFS respondents in Northern Ireland).

This work represents "proof of concept" research and the results should not be treated as revised official statistics from the LFS. Users of our labour market statistics should continue to use the routinely published data, and LFS microdata users should continue to use the datasets currently available via the Secure Research Service and other outlets. However, our research to date supports our existing [user guidance](#) that PAYE RTI currently provides the most reliable measure of employees, particularly during the period covered by this analysis. We will use feedback on this research from our users when considering the potential use of this approach in the production of our labour market statistics.

Our analysis suggests patterns in apparent under- or over-representation of PAYE employees in the LFS respondent sample have varied over time. There was apparent under-representation of PAYE employees in 2019; changing to apparent over-representation during the peak of the coronavirus (COVID-19) pandemic from mid-2020 to the end of 2021; returning to apparent under-representation from mid-2022 to mid-2024; and little evidence of under- or over-representation in 2025.

Accordingly, recalibrating the LFS survey weights to PAYE totals using the approach outlined in this article increases the estimated employment rate in 2019; decreases it from mid-2020 to the end of 2021; increases it from mid-2022 to mid-2024; and has little impact in 2025. The impact of recalibration on estimated unemployment and economic inactivity rates mirrors that for the employment rate.

This experimental analysis demonstrates the potential value of using administrative data in the estimation process for labour market surveys. Further statistical research and feasibility testing, including conducting the same analysis on the Transformed Labour Force Survey (TLFS), is needed before we can determine whether this methodology should be incorporated into the production of official labour market statistics.

These findings are subject to change when the LFS data are reweighted to reflect the latest available population estimates and other improvements.

2 . The Labour Force Survey

Survey methodology

The Labour Force Survey (LFS) provides timely information on labour market conditions across the UK, including indicators such as rates of employment, unemployment and economic inactivity (see [Section 6: Definitions](#)). Its primary purpose is to inform economic policy, as well as providing insights to businesses and the general public.

The LFS's target population comprises individuals living in private households, as well as those in a limited set of communal establishments: student halls of residence and National Health Service (NHS) staff accommodation. Each calendar quarter, a systematic random sample is drawn from national address lists ordered by postcode; in the case of multiple households living at the same address, one household is selected. All residents within sampled households are eligible to participate in the survey. Once a household is selected into the sample, they are initially interviewed and then followed up for a further four consecutive quarters.

As is standard practice for household surveys, the LFS sample is weighted to ensure the resulting statistical estimates are representative of the target population using a three-step process:

- design weights are calculated as the inverse of the probability of a household being selected into the sample
- the design weights are adjusted for area-based factors related to the probability of a household not responding to the survey given they were selected into the sample: region, relative area deprivation, and Output Area classification
- the non-response-adjusted design weights are calibrated such that the population group totals in the weighted sample (that is, the sum of the weights within groups) match those in the target population according to several characteristics, or auxiliary variables

The calibration stage requires the presence of one or more auxiliary variables that meet the following criteria:

- highly correlated with the survey variables of interest
- available for both sample respondents and in the population of interest
- measured on a like-for-like basis between the sample and population datasets

Calibration aims to minimise changes to the non-response-adjusted design weights while preserving population totals across the following partitions:

- Partition 1: Local authority district of residence
- Partition 2: Sex (men, women) by age group (0 to 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 years and over for Great Britain; 0 to 15, 16 to 17, 18 to 20, 21 to 22, 23 to 24, 25 years and over for Northern Ireland)
- Partition 3: Sex (men, women) by country or region of residence (Wales, Scotland, Northern Ireland, the nine former Government Office Regions (GORs) of England) by age group (0 to 15, 16 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, 65 to 74, 75 years and over)
- Partition 4: Country of residence (Great Britain, Northern Ireland) by sex (men, women) by age group (0 to 4, 5 to 9, 10 to 15, 16 to 19, 20 to 24, 25 to 29, 30 to 34, 35 to 39, 40 to 44, 45 to 49, 50 to 54, 55 to 59, 60 to 64, 65 to 69, 70 to 74, 75 years and over)
- Partition 5: Housing tenure group (owned outright, mortgage or loan, rent)

Both the non-response adjustment to the design weights and the calibration stage mitigate against non-response bias; that is, systematic over- or under-estimation of labour market quantities of interest (such as employment rates) caused by certain groups being less likely to respond to the survey than others. This is done by controlling for factors that are likely to be related to both the likelihood of response and the quantity of interest. However, while this approach may reduce non-response bias, it will not eliminate it if the likelihood of response is directly affected by employment status (because factors such as age, sex and geography are imperfectly correlated with employment status).

Further information on the survey methodology can be found in our [LFS quality and methodology information article](#) and [LFS user guidance](#). The LFS is due to be replaced by the Transformed Labour Force Survey (TLFS) in the near future, as described in our latest [labour market transformation update on progress and plans](#). However, the LFS remains the focus of our current "proof of concept" research into using administrative data in labour market statistics. Similar analysis will be conducted on the TLFS once the survey data and methods have matured.

Falling response rates and potential for bias

Response rates have fallen from around 80% of sampled households in the early 1990s to under 25% by the end of 2025 (Figure 1).

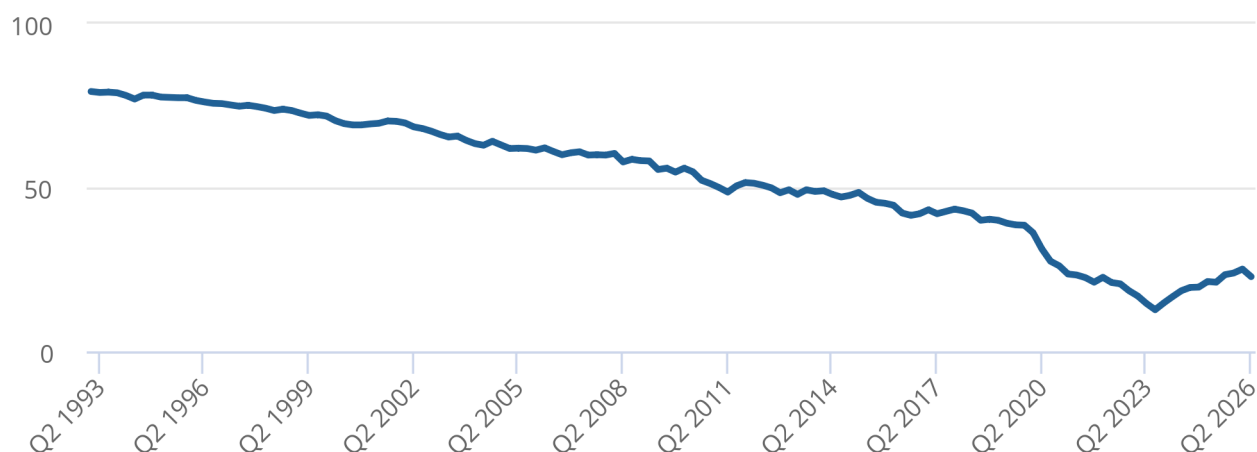
The coronavirus (COVID-19) pandemic exacerbated the decline, with household response rates in Great Britain falling from 38.5% in Quarter 4 (Oct to Dec) 2019 to a low of just 12.7% in Quarter 3 (July to Sept) 2023. As part of work to address quality concerns with the LFS, response rates have shown clear improvement since then, and the achieved sample sizes are now close to their pre-pandemic level. However, the decline in response rates reflects a long-term trend rather than an artefact of the pandemic.

Figure 1: LFS response rates have been in decline over several decades

Labour Force Survey (LFS) household response rates, Quarter 1 (Mar to May) 1993 to Quarter 2 (Apr to June) 2026, Great Britain

Figure 1: LFS response rates have been in decline over several decades

Labour Force Survey (LFS) household response rates, Quarter 1 (Mar to May) 1993 to Quarter 2 (Apr to June) 2026, Great Britain



Source: Labour Force Survey from the Office for National Statistics

Notes:

1. From 1993 to 2005: Quarter 1 (Mar to May), Quarter 2 (June to Aug), Quarter 3 (Sept to Nov) and Quarter 4 (Dec to Feb).
2. From 2006 onwards: Quarter 1 (Jan to Mar), Quarter 2 (Apr to June), Quarter 3 (Jul to Sept) and Quarter 4 (Oct to Dec).
3. Response rates from Quarter 1 1993 to Q1 2001 are taken from an article published in the [Economic and Labour Market Review](#).
4. Response rates from Quarter 2 2001 onwards are taken from our [Labour Force Survey performance and quality monitoring reports](#).
5. Response rates relate to all five survey waves combined and exclude imputed households.

While falling response rates do not necessarily mean more non-response bias, they do at least suggest more potential for any bias to materially affect the survey estimates. This would occur if individuals who do not respond to the LFS tend to differ from responding individuals in ways related to the labour market quantities of interest, which is not accounted for by the existing weighting variables. For example, estimates of the employment rate will be biased if employed people are more or less likely to respond to the LFS than non-employed people, after accounting for their age, sex, geography of residence and housing tenure status.

Using administrative data to identify and adjust for non-response bias

Administrative data are produced when citizens or organisations interact with a public service, and they are often the byproduct of a service being administered rather than data being collected for statistical purposes.

The Office for National Statistics (ONS) receives a monthly data extract from the Pay As You Earn (PAYE) Real Time Information (RTI) system, the primary purpose of which is the administration of Income Tax and National Insurance contributions by HM Revenue and Customs (HMRC). As well as individual-level information on gross pay, tax paid and other deductions, the dataset contains basic demographic variables such as age, sex, and area of residence (all based on information supplied by employers).

Administrative datasets may have potential for identifying and adjusting for non-response bias in statistical estimates from survey samples if they:

- are available for everyone in the population of interest (both survey respondents and non-respondents)
- contain variables that are highly correlated with the survey quantity of interest

The PAYE dataset meets both of these criteria. First, the data are available for nearly all payrolled employees registered in UK-based PAYE schemes. Second, employee status derived from PAYE records is highly correlated with labour market variables collected on the LFS. After linking LFS responses to PAYE records from Quarter 1 (Jan to Mar) 2019 to Quarter 4 (Oct to Dec) 2025 (see [Section 3: Methodology for incorporating administrative data into survey estimation](#)), the average Phi correlation coefficients were 0.88 for employee status, 0.78 for employment status, and 0.76 for economic activity status (see [Section 6: Definitions](#)). Quarterly correlation coefficients over the study period can be found in the [accompanying dataset](#). However, the degree of correlation between PAYE-derived employee status and LFS unemployment status is lower, with an average Phi coefficient of negative 0.30 among the economically active.

Research into using administrative data related to the labour market (including employment registers, and tax and benefits databases) for non-response bias identification and weighting has been conducted in several countries including [Norway](#), [Austria](#), [Germany](#) and the [United States of America](#). Integrated use of auxiliary information (including administrative data) to mitigate against non-response bias is also recommended by the United Nations Economic Commission for Europe (UNECE) in their [Practical guide to cost-effectiveness and non-response treatment for social surveys in national statistical organisations](#) (PDF, 4.4MB). We therefore sought to investigate whether PAYE records could improve the quality of estimates from the LFS through their use in calibration.

3 . Methodology for incorporating administrative data into survey estimation

Our proposed methodology involves comparing the number of employees estimated from the weighted Labour Force Survey (LFS) sample with the number in the population according to Pay As You Earn (PAYE) administrative data. Differences between these employee counts may shed light on any employment-related non-response in the LFS; that is, the extent to which working people are more or less likely to be represented in the LFS respondent sample than non-working people after applying the current survey weights.

However, differences between employee counts estimated from the weighted LFS sample and in the PAYE population may only be attributed to non-response if we first rule out other sources of systematic difference between the two time series:

- differences in the definition of an employee
- differences in population coverage

Systematic differences between how individuals self-report their employee status on the LFS versus how it is inferred from PAYE records may arise through:

- the purpose of data collection (LFS data are collected for statistical reasons and conform to International Labour Organization (ILO) definitions, whereas PAYE records are an administrative byproduct of the Income Tax and National Insurance system)
- timing effects (individual LFS responses relate to a single reference week within a month, whereas PAYE records provide an indication of whether an individual was ever an employee at any point during the month)
- source-specific measurement errors (for example, recall or proxy-respondent error in the LFS, and imperfect monthly calendarisation of PAYE records)

To overcome these differences, we link individual LFS responses to the PAYE dataset, enabling us to use a consistent PAYE-based employee definition for both LFS respondents and the PAYE population. Although the PAYE-based employee estimate is an imperfect measure of the ultimate variable of interest (ILO employment), the two are highly correlated. Therefore, differences between LFS-derived and PAYE-derived estimates of PAYE employee counts are likely to be informative in terms of the size and direction of any employment-related sample misrepresentation.

The main sources of over-coverage in the PAYE dataset relative to the LFS target population are individuals living abroad but taxed via UK-based PAYE schemes; and employees living in communal establishments other than National Health Service (NHS) staff accommodation and student halls of residence (such as care homes, hotels, prisons and military bases).

This over-coverage may lead us to conclude that PAYE employees are under-represented in the LFS sample, even if there is no employment-related non-response bias or the bias is in the other direction. To overcome this, we downscale the population-level PAYE employee totals by a factor calculated from other data sources.

Once the employee counts estimated from the weighted LFS sample and in the PAYE population have been made consistent in terms of definitions and coverage, differences between the employee time series may provide insights into the size and direction of any employment-related non-response in the LFS. It is also possible to reweight the LFS using the PAYE data, allowing us to assess the impact of non-response bias reduction on headline estimates of employment, unemployment and economic inactivity rates.

The remainder of this section explains the detailed steps of our proposed methodology.

Step 1: prepare the PAYE dataset and link to LFS responses

Pension payments are removed from the PAYE dataset, leaving payments made by employers to employees. However, each record on the dataset represents a payment made on a given date, rather than the period of employment when the work was undertaken by the employee. The latter is not stored in the PAYE record and must be estimated via a calendarisation process, as explained in our [user guide to earnings and employment from PAYE RTI](#).

The Office for National Statistics's (ONS's) Demographic Index (DI), as outlined on our [Reference Data Management Framework](#) webpage, enables record-level linkage between the LFS and PAYE datasets. This means that, within a given time period, each LFS respondent can be matched to their PAYE information, if it exists.

Between 2019 and 2024, the percentage of LFS respondents in Great Britain who could be indexed to the DI varied from 96.2% to 97.7%. Through clerical review of a random sample of approximately 10,000 automatically indexed LFS respondents, the estimated precision (the proportion of links deemed to be correct) ranged from 99.5% to 99.7% across the six years. Clerical review of a random sample of approximately 4,500 individuals produced an estimated recall (the proportion of matches deemed to be true that were automatically linked) ranging from 99.2% to 99.6%.

Our exploratory analysis uses historical data to the end of 2025, allowing us to link LFS responses to PAYE records in the same month. In practice, if our proposed methodology was to be implemented in the production of monthly labour market statistics, we would need to consider whether calendarised, DI-indexed PAYE records could be made available in time to use in the latest reference period each month.

Previously supplied PAYE data are also revised each month when the latest extract is supplied by HM Revenue and Customs (HMRC) to the ONS, reflecting additional PAYE submissions received. If PAYE data timeliness and revisions are found to represent a practical challenge to implementation, LFS responses may need to be linked to lagged PAYE information (for example, from the month or quarter prior to the reference period).

Step 2: derive PAYE employee status for LFS respondents

LFS respondents who can be linked to a PAYE record in the same month are classified as PAYE employees in that month (PAYE calibration variable equals 1).

LFS respondents who cannot be linked to a PAYE record in the same month are classified as PAYE non-employees in that month (PAYE calibration variable equals 0).

As 2% to 4% of LFS respondents cannot be indexed to the DI, it is impossible to infer whether these individuals are PAYE employees. For these cases, the PAYE employee status is imputed using the period-specific probability of being a PAYE employee given the ILO employee status, computed among individuals who could be indexed to the DI. We assume this probability would be the same for the non-indexed cases.

For example, if 98% of ILO employees are also PAYE employees among DI-indexed LFS respondents, the PAYE calibration variable will be set to 0.98 for ILO employees who cannot be indexed to the DI. Conversely, if 5% of ILO non-employees are PAYE employees among DI-indexed LFS respondents, the PAYE calibration variable will be set to 0.05 for ILO non-employees who cannot be indexed to the DI. Thus, in this example, the PAYE calibration variable will be equal to 0 or 1 for the 96% to 98% of LFS respondents who can be indexed to the DI, and 0.05 or 0.98 for the 2% to 4% of respondents who cannot be indexed to the DI.

Step 3: adjust population PAYE employee totals for over-coverage relative to the LFS target population

Ad hoc analysis of self-assessment data performed by HMRC indicates that approximately 0.12% of PAYE employees in a given financial year are resident outside the UK, and are therefore not within the LFS target population. This figure has remained stable since the coronavirus (COVID-19) pandemic.

Among employees in the PAYE dataset who were aged 16 years and over, resident in England or Wales in March 2021, and who could be linked to Census 2021 via the DI, 0.11% were living in communal establishments other than NHS staff accommodation and student halls of residence according to the census. Table 1 shows this percentage broken down by age group and sex.

Table 1: Percentage of people in communal establishments, other than NHS staff accommodation and student halls of residence, among PAYE employees linked to Census 2021, England and Wales, March 2021

Age range	Men	Women	Both sexes
16 to 24 years	0.34	0.15	0.24
25 to 34 years	0.21	0.09	0.15
35 to 49 years	0.10	0.04	0.07
50 to 64 years	0.06	0.04	0.05
65 years and over	0.18	0.38	0.28
16 years and over	0.15	0.08	0.11

Source: Pay As You Earn (PAYE) Real Time Information from HM Revenue and Customs, Census 2021 for England and Wales from the Office for National Statistics

The population-level PAYE employee totals were downscaled to account for the overall extent of over-coverage relative to the LFS target population, calculated as the product of the non-UK residency and age-sex-specific communal establishment residency prevalence estimates given previously. The downscaling factor was assumed to be constant over time.

There are also some employees who are within the scope of the LFS target population but are not included in the PAYE dataset:

- employees belonging to a PAYE scheme in which all employees are below the National Insurance Lower Earnings Limit (currently £559 per month)
- employees whose [employer is exempt from filing their PAYE returns digitally](#), as detailed by HMRC guidance
- employees taking an extended period of unpaid leave (such as those on long-term sickness absence or a career break)
- employees removed from the dataset by HMRC for reasons related to data protection or propriety

These exclusions do not represent under-coverage in the PAYE dataset relative to the LFS target population. These individuals are still in the population, but they would be classified as a PAYE non-employee rather than a PAYE employee (in both the linked LFS sample and in the population). This is accounted for in Step 2 (derive PAYE employee status for LFS respondents to align employee definitions between the data sources), so no under-coverage adjustment to population-level PAYE employee totals is required.

The PAYE dataset does not contain information for people who are solely self-employed and do not pay themselves via PAYE, so our proposed methodology involves comparing sample and population counts of employees rather than employment. This is not a problem for the methodology as PAYE employee status is highly correlated with the survey variables of interest, suggesting that it is a suitable auxiliary variable with which employment, unemployment and economic inactivity rates can be estimated. However, the methodology could be improved if timely administrative data on the self-employed were available ([see Section 5: Further work](#) for more details).

Step 4: reweight the LFS using PAYE data

The LFS sample can be reweighted to PAYE employee totals by adding a sixth calibration partition to the five partitions currently used in the LFS calibration (see [Section 2: The Labour Force Survey](#)):

- Partition 6: PAYE employee status (employee, non-employee) by sex (men, women) by age group (16 to 24, 25 to 34, 35 to 49, 50 to 64, 65 years and over)

If PAYE employees were over-represented in the original weighted sample, recalibrating to PAYE totals will downscale the weights of employees and, correspondingly, upscale the weights of non-employees (as the total population size is held constant). Conversely, if PAYE employees were under-represented in the original weighted sample, recalibrating to PAYE totals will upscale the weights of employees and downscale the weights of non-employees.

This approach of recalibrating the LFS weights to PAYE totals ensures the original calibration constraints (LFS population totals) are maintained, preserving any reduction in non-response bias achieved by the original calibration that is independent of PAYE employee status. It also maintains coherence between ONS data sources, which often share a common total population size (summed weights) by factors such as age and sex.

Disaggregating PAYE employee status by age group and sex is enabled by age and sex information being available on the population PAYE dataset. This disaggregation is important because the extent and direction of any over- or under-representation of PAYE employees in the LFS sample may differ by demographic characteristics, which may also be related to the survey variables of interest.

The PAYE dataset provides information on the number of PAYE employees in the population in a given month, but not the number of non-employees. Therefore, the total population size (by age group and sex) is held constant before and after recalibration, and the number of PAYE non-employees in the population (by age group and sex) is then inferred as the difference between the total population size and the number of PAYE employees.

4 . Results of applying the methodology

Implied representativeness of PAYE employees in the LFS sample

Figure 2 shows the number of employees in Great Britain estimated from the weighted Labour Force Survey (LFS) sample using self-reported employee status (as per International Labour Organization (ILO) definitions) and linked Pay As You Earn (PAYE) records (as per administrative definitions). Both time series can be compared with the number of employees estimated from the population-level PAYE dataset (adjusted for over-coverage relative to the LFS target population).

Employee counts estimated from the population-level PAYE dataset are consistently higher than those estimated from the LFS using self-reported (ILO) employee information. However, it is not possible to infer the extent to which this discrepancy is attributable to under-representation of employees in the weighted LFS sample, because definitional differences between the time series also exist.

Conversely, PAYE employee counts estimated from the linked LFS sample are comparable with those estimated from the population-level PAYE dataset because both are based on the same employee definition, as described in [Section 3: Methodology for incorporating administrative data into survey estimation](#). Though some differences between the series remain, the sample-based estimates using the PAYE-based employee definition are markedly closer to the population-level PAYE estimates than when the self-reported (ILO) employee definition is used. This emphasises the importance of aligning employee definitions between the data sources before making comparisons between the time series.

Differences between LFS-derived and population PAYE-derived estimates of PAYE employee counts give an indication of the apparent size and direction of any over- or under-representation of PAYE employees in the LFS respondent sample that remains after applying the current survey weights. This may at least partly translate to non-response bias in rates of employment, unemployment and economic inactivity estimated from the LFS.

However, these differences can only be interpreted as fully representing non-response bias under the assumption that there are no other systematic differences between the time series. While we have accounted for differences in employee definitions and population coverage between the data sources, we cannot rule out that other differences may exist.

For example, the population totals to which the LFS is weighted are estimates rather than true values, as described in our [Impact of reweighting on Labour Force Survey key indicators: December 2024 article](#). The main sources of uncertainty include the size of the starting population (estimated from the last decennial census) and estimated net migration in subsequent years. If employees tend to be under-estimated (or over-estimated) in the population totals but are accurately captured in (or omitted from) the PAYE data, this will induce a difference between LFS-derived and PAYE-derived estimates of PAYE employee counts, even if non-response bias is not present.

Figure 2: Patterns in the implied representativeness of PAYE employees in the weighted LFS respondent sample have varied over time

Number of employees aged 16 years and over estimated from weighted Labour Force Survey (LFS) responses and linked Pay As You Earn (PAYE) records, and from population PAYE data, Quarter 1 (Jan to Mar) 2019 to Quarter 4 (Oct to Dec) 2025, Great Britain

Notes

1. Y-axis scale does not start at 0.
2. Estimates are not seasonally adjusted.
3. LFS employee estimates are based on whether respondents reported being an employee in their reference week.
4. PAYE employee estimates from the LFS sample are based on whether LFS respondents were PAYE employees in the same month as their LFS reference week.
5. Population-level PAYE employee estimates relate to whether individuals were employees at any time throughout each month, averaged across the three months within each calendar quarter. This is different to the PAYE employee estimates published in our monthly Labour market overview, UK bulletin, which reflect the proportion of each month that individuals spent as employees.
6. There are no confidence intervals around the population PAYE estimates as there is no sampling error. Other sources of uncertainty, such as measurement error, exist but cannot easily be quantified.
7. The variance estimation methodology used for this analysis is very similar, though not identical to the methodology used to compute confidence intervals in our routinely published labour market statistics.
8. Population PAYE employee estimates are adjusted to account for over-coverage relative to the LFS target population.
9. Sample-based estimates from LFS responses and linked PAYE records are weighted using the current LFS weights.

Download

After accounting for differences in employee definitions (via linkage) and population coverage (via adjustment), the number of PAYE employees in Great Britain estimated from the weighted LFS sample is less than the number in the population estimated from administrative data throughout 2019, by an average of 217,000 employees per quarter.

Lack of availability and time during the working day may have contributed to employees being less able to respond to the survey than non-employees, particularly for individuals who work away from their home.

This pattern changes during the peak of the coronavirus (COVID-19) pandemic, from Quarter 2 (Apr to June) 2020 to the end of 2021, when the number of PAYE employees estimated from the weighted LFS sample is greater than the number in the population. Over this period, the average difference between the estimates is 332,000 employees per quarter, with a maximum difference of 644,000 employees in Quarter 1 (Jan to Mar) 2021.

This may reflect pandemic-related social restrictions affecting LFS data collection, most notably widespread homeworking and the temporary replacement of face-to-face interviews with telephone interviews for Wave 1 LFS respondents. This may have made it easier for employed people to respond to the survey compared with pre-pandemic.

The pre-pandemic pattern returns from the start of 2022 to the end of 2024 but to a greater extent. During this period, the number of PAYE employees estimated from the weighted LFS sample is less than the number in the population by an average of 365,000 employees per quarter.

This coincides with the easing of pandemic-related social restrictions (the UK government's [Living with COVID-19](#) guidelines were published in February 2022), potentially shifting the composition of the respondent sample away from employed people. This was also a period of declining and historically low levels of LFS response, falling to approximately 14,000 household interviews achieved and a less than 13% response rate in Great Britain in Quarter 3 (July to Sept) 2023.

Our analysis suggests little evidence of non-response bias in 2025, with the number of PAYE employees estimated from the weighted LFS sample being lower than the number in the population by just 10,000 employees per quarter on average.

Estimates throughout 2025 include the effect of improvements in LFS data collection and sampling methods aiming to increase achieved sample sizes (which had almost returned to their pre-pandemic level by the end of 2025) and response rates. These include reinstating the sample boost, increased response-chasing activities, recruitment of additional field interviewers, and financial incentives for responding that are substantially higher than they were before and during the pandemic. Further information can be found in our [Labour Force Survey: planned improvements and its reintroduction article](#) and [Labour market transformation - update on progress and plans: July 2025 article](#).

Taken together, these factors may explain the limited evidence of employment-related non-response bias in 2025, rather than a return to the 2019 (pre-pandemic) situation of employees appearing to be less likely to respond to the LFS than non-employees.

LFS-derived and PAYE-derived estimates of PAYE employee counts broken down by age group and sex can be found in the [accompanying dataset](#). LFS-derived estimates of PAYE employee counts are greater than the population PAYE-derived estimates during the peak pandemic period for people aged 25 to 49 years and women. However, over the period following the peak pandemic, the biggest negative differences between the LFS-derived and PAYE-derived estimates of PAYE employee counts can be seen in people aged 50 to 64 years and men.

Impact on labour market estimates of using PAYE records to reweight the LFS

Calibrating the LFS survey weights to PAYE employee totals using the approach described in this article results in an employment rate in Great Britain that is lower than that estimated using the current LFS weights during the peak pandemic period, from Quarter 2 2020 to the end of 2021 (Figure 3).

During this period, the employment rate is on average 0.57 percentage points lower per quarter when the LFS is reweighted to PAYE totals, equating to an average of 230,000 fewer people aged 16 to 64 years estimated to be in employment each quarter.

This pattern reverses at the start of 2022 until the end of 2024, when the employment rate is an average of 0.42 percentage points higher per quarter when the LFS is reweighted to PAYE totals. This equates to an average of 174,000 more people estimated to be in employment each quarter.

The original and recalibrated estimates are closer in 2025, with an average reduction in the employment rate of just 0.08 percentage points per quarter. This equates to an average of 33,000 fewer people estimated to be in employment each quarter using the recalibrated weights.

Differences between the original and recalibrated series provide an indication of apparent non-response bias in the original series. However, both series are estimates and the true employment rate, and therefore the true extent of non-response bias, remain unknown. It is also possible that residual non-response bias remains in the recalibrated estimate because Pay As You Earn (PAYE) employee status is imperfectly correlated with Labour Force Survey (LFS) employment status.

Figure 3: The effect on the employment rate of calibrating the LFS to PAYE information varies over time

Estimated rate of employment among people aged 16 to 64 years, before and after recalibration of Labour Force Survey (LFS) weights to Pay As You Earn (PAYE) employee totals, Quarter 1 (Jan to Mar) 2019 to Quarter 4 (Oct to Dec) 2025, Great Britain

Notes

1. Y-axis scale does not start at 0.
2. Estimates are not seasonally adjusted.
3. The variance estimation methodology used for this analysis is very similar, though not identical to the methodology used to compute confidence intervals in our routinely published labour market statistics.

Changes in the estimated rate of employment are largely mirrored by opposing changes in the estimated rates of economic inactivity (Figure 4) and unemployment (Figure 5).

After calibrating the LFS survey weights to PAYE employee totals, the inactivity and unemployment rates in Great Britain are higher than those estimated using the current LFS weights during the peak pandemic period, from Quarter 2 2020 to the end of 2021. During this period, the inactivity and unemployment rates are on average 0.49 and 0.12 percentage points higher per quarter, respectively, when the LFS is reweighted to PAYE totals.

This pattern reverses at the start of 2022 until the end of 2024, when the inactivity and unemployment rates are on average 0.37 and 0.09 percentage points lower per quarter, respectively, when the LFS is reweighted to PAYE totals.

The original and recalibrated estimates are closer in 2025, with average differences in the inactivity and unemployment rates of just 0.06 and 0.02 percentage points per quarter, respectively.

Pay As You Earn (PAYE) employee status and Labour Force Survey (LFS) unemployment status are only weakly correlated. Therefore, small differences between the original and recalibrated estimates do not necessarily indicate an absence of non-response bias in terms of unemployment status.

It should be noted that absolute differences between the original and recalibrated estimates may generally be smaller for unemployment rates than for inactivity rates partly because the unemployment rate is lower in absolute terms. However, relative differences between the original and recalibrated estimates are generally bigger for unemployment rates than for inactivity rates.

For example, in Quarter 1 2021 (the period of biggest positive changes in rates between the original and recalibrated estimates), recalibration leads to a 0.62 percentage point increase in the estimated inactivity rate and a smaller 0.25 percentage point increase in the unemployment rate. However, in relative terms, the recalibrated unemployment rate is 5.16% higher than the original estimate, compared with a 1.69% increase for the inactivity rate.

Conversely, in Quarter 1 2024 (the period of biggest negative changes in rates between the original and recalibrated estimates), recalibration leads to a 0.64 percentage point decrease in the estimated inactivity rate and a smaller 0.19 percentage point decrease in the unemployment rate. However, in relative terms, the recalibrated unemployment rate is 4.34% lower than the original estimate, compared with a 1.72% decrease for the inactivity rate.

Estimated rates and levels of employment, economic activity and inactivity, unemployment, employees and self-employment broken down by age group, sex and country of residence can be found in the [accompanying dataset](#).

Figure 4: The effect of calibrating the LFS to PAYE information on estimates of economic inactivity largely mirrors the effect on estimates of employment

Estimated rate of economic inactivity among people aged 16 to 64 years, before and after recalibration of Labour Force Survey (LFS) weights to Paye As You Earn (PAYE) employee totals, Quarter 1 (Jan to Mar) 2019 to Quarter 4 (Oct to Dec) 2025, Great Britain

Notes

1. Y-axis scale does not start at 0.
2. Estimates are not seasonally adjusted.
3. The variance estimation methodology used for this analysis is very similar, though not identical to the methodology used to compute confidence intervals in our routinely published labour market statistics.

Figure 5: The effect of calibrating the LFS to PAYE information is smaller for unemployment than inactivity in absolute terms, but bigger in relative terms

Estimated rate of unemployment among people aged 16 years and over, before and after recalibration of Labour Force Survey (LFS) weights to Pay As You Earn (PAYE) employee totals, Quarter 1 (Jan to Mar) 2019 to Quarter 4 (Oct to Dec) 2025, Great Britain

Notes

1. Y-axis scale does not start at 0.
2. Estimates are not seasonally adjusted.
3. The variance estimation methodology used for this analysis is very similar, though not identical to the methodology used to compute confidence intervals in our routinely published labour market statistics.

Download

Impact on uncertainty estimates of using PAYE records to reweight the LFS

As well as reducing apparent non-response bias in estimated employment and economic inactivity rates, calibrating the LFS survey weights to PAYE employee totals also has the potential to reduce [statistical uncertainty](#) for these variables. This is because PAYE employee status is highly correlated with the survey variables of interest, so a large proportion of the variance in these variables can be explained by the variance in PAYE employee status. This variance-reducing effect is generally larger than the variance-increasing effect of increasing the complexity of the calibration and therefore the variability in the survey weights.

The reduction in uncertainty after recalibrating the LFS weights to PAYE totals can be seen in Figures 3 and 4 as a narrowing of the [95% confidence intervals](#) around the estimated rates of employment and economic inactivity. For example, in Quarter 4 (Oct to Dec) 2025, the confidence interval for the employment rate ranges from 74.7% to 75.6% before recalibrating to PAYE totals, a coverage in possible values of 0.88 percentage points. This coverage falls to 0.63 percentage points after recalibration, a reduction of 0.25 percentage points (or 28.2% in relative terms).

There are also potential gains in precision for more granular estimates, at least by age group and sex (which are both included in the additional PAYE-based calibration partition). For example, when the employment rate in Quarter 4 2025 is broken down by age group, the coverage of the confidence interval decreases by 0.78 percentage points (29.0%) for people aged 16 to 24 years, 0.55 percentage points (29.9%) for those aged 25 to 34 years, 0.34 percentage points (26.3%) for those aged 35 to 49 years, and 0.43 percentage points (27.8%) for those aged 50 to 64 years.

The reduction in uncertainty can also be quantified using [coefficients of variation](#), which account for both the variance of an estimate and the magnitude of the estimate itself. After calibrating to PAYE totals, the coefficients of variation fall in relative terms by between 26.4% and 33.7% for the employment rate (Figure 6) and between 20.3% and 26.5% for the inactivity rate, indicating an improvement in precision (or a reduction in uncertainty). Decreases in the coefficient of variation after recalibration are smaller for the unemployment rate than for the employment and economic inactivity rates, with relative reductions ranging from 1.6% to 4.5%.

Coefficients of variation for estimated rates and levels of employment, economic activity and inactivity, unemployment, employees and self-employment broken down by age group, sex and country of residence can be found in the [accompanying dataset](#).

Figure 6: Calibrating the LFS to PAYE information reduces statistical uncertainty around the estimated employment rate

Coefficient of variation for the rate of employment among people aged 16 to 64 years, before and after recalibration of Labour Force Survey (LFS) weights to Paye As You Earn (PAYE) employee totals, Quarter 1 (Jan to Mar) 2019 to Quarter 4 (Oct to Dec) 2025, Great Britain

Notes

1. The variance estimation methodology used for this analysis is very similar, though not identical to the methodology used to compute confidence intervals in our routinely published labour market statistics.

5 . Further work

This experimental analysis demonstrates the potential value of using HM Revenue and Customs (HMRC) Pay As You Earn (PAYE) administrative data in the estimation process for labour market surveys. However, this analysis represents "proof of concept" research and further development work is required, so the results should not be treated as revised official statistics from the Labour Force Survey (LFS).

Users of our labour market statistics should continue to use the routinely published data, and LFS microdata users should continue to use the datasets currently available via the Secure Research Service and other outlets. However, our research to date supports our existing [user guidance](#) that PAYE Real Time Information (RTI) currently provides the most reliable measure of employees, particularly during the period covered by this analysis.

We welcome feedback from users to help inform the potential use of this approach in the production of our labour market statistics. Please contact us via email at labour.market.transformation@ons.gov.uk to share your views.

Our experimental analysis to date has focussed on the application of the proposed methodology to the LFS, but this is due to be replaced by the Transformed Labour Force Survey (TLFS) in the near future. The proposed approach may also be applicable to the TLFS, though an impact assessment can only be conducted once the underlying survey methodology of the TLFS is sufficiently mature and an adequate time series is available for analysis. We will also use the proposed methodology to help us understand differences between the LFS, TLFS and PAYE time series as the TLFS methodology is refined and we approach transition from the LFS to the TLFS. If the methodology demonstrates benefit and feasibility for the TLFS, implementation may take place after the initial transition. We will update users on our plans at a later date.

Before deciding whether to incorporate administrative data into the methodology for producing official labour market statistics, as explored in this article, additional statistical research and feasibility testing is required. More information on this is included in the remainder of this section.

Re-run the analysis after the LFS is reweighted

This reweighting will reflect the latest available population estimates, as well as other improvements.

Test the methodology on Northern Ireland data

The analysis needs to be extended from Great Britain to full UK-wide coverage when the data linkage allows.

Investigate PAYE data timeliness and revisions

To operationalise the proposed methodology, we need to consider whether processed PAYE records could be made available in time for the latest reference period each month, and whether the initial monthly PAYE extract, which is revised in later months, is sufficiently informative. If not, we will assess the impact of calibrating to lagged PAYE information.

Assess the impact on LFS-derived labour market indicators other than employment, unemployment and inactivity

We want to assess, for example, average working hours and earnings, which may be less well correlated with PAYE employee status than employment, unemployment and economic inactivity statuses.

Extend the impact assessment to rates of period-on-period growth in headline labour market measures

This assessment includes variance-of-change estimation for the recalibrated time series.

Investigate calibrating to the PAYE earnings distribution

The pay information on the PAYE dataset is not considered by the current proposed methodology, but may be beneficial, particularly for earnings estimates, if LFS non-response is related to employee earnings as well as employee status.

Test sensitivity to linkage error

We need to rule out the possibility that differences between LFS-derived and population PAYE-derived estimates of PAYE employee counts are driven by patterns in linkage error rather than non-response. However, linkage of LFS respondents to the Demographic Index is performed with high precision (99.5% or above), so we expect the impact of linkage error to be small.

Incorporate data for the self-employed

Supplementing PAYE records, which cover only employees, with HMRC self-assessment records, which cover the self-employed, could further improve the estimation. In practice, the annual self-assessment data will never be sufficiently timely to use in routine production of monthly labour market statistics, though this could be improved in the future by considering quarterly self-employment returns from the [Making Tax Digital for Income Tax](#) system. Nonetheless, we intend to perform a retrospective analysis using historical records to assess the sensitivity of our current findings to including data on the self-employed.

6 . Definitions

Economic inactivity

The number of people without a job who have not actively sought work in the last four weeks, and/or are not available to start work in the next two weeks.

Economic inactivity rate

Calculated by dividing the inactivity level among people aged 16 to 64 years by the total population size in that age group.

Employees

People who are in employment and paid a wage by an employer for the work they do.

Employment

The number of people who did paid work (as an employee or self-employed), those who had a job that they were temporarily away from, those placed with employers on government-supported training and employment programmes, and those doing unpaid family work. Note that an individual's employment status according to this statistical definition may differ from that as defined in [employment law and for tax purposes](#).

Employment rate

Calculated by dividing the employment level among people aged 16 to 64 years by the total population size in that age group.

Self-employed

People who work on their own account, whether or not they have employees.

Unemployment

The number of people without a job and have actively sought work in the last four weeks and are available to start work in the next two weeks; or who are out of work but have found a job and are waiting to start it in the next two weeks.

Unemployment rate

Calculated by dividing the unemployment level among people aged 16 years and over by the total number of economically active people (those in the population who are not economically inactive) in that age group.

7 . Data on investigating non-response to the Labour Force Survey using administrative data

[Investigating non-response to the Labour Force Survey using administrative data, Great Britain](#)

Dataset | Released 28 September 2026

Using Pay As You Earn data to identify and adjust for potential employment-related non-response to the Labour Force Survey. These are not official statistics.

8 . Related links

[Labour market overview, UK: September 2026](#)

Statistical bulletin | Released 15 September 2026

Estimates of employment, unemployment, economic inactivity, and other employment-related statistics for the UK.

[Labour Force Survey - user guidance](#)

Methodology | Last revised 15 September 2026

Guidance about the background and methodology of the Labour Force Survey (LFS).

[User guide to earnings and employment from Pay As You Earn Real Time Information](#)

Methodology | Last revised 25 July 2025

A comprehensive overview of the background and methodology of the Pay As You Earn (PAYE) Real Time Information (RTI) statistics.

[A guide to labour market statistics](#)

Methodology | Last revised 16 June 2020

Explanation of the major concepts that exist within the labour market and their relationship to each other.

9 . Cite this page

Office for National Statistics (ONS), released 28 September 2026, ONS website, supporting methodology article, [Investigating non-response to the Labour Force Survey using administrative data](#)