

Exploring the use of external data to assess for observed bias in Labour Force Survey estimates: interim findings

A report on progress to assess potential bias in the Labour Force Survey through a comparison against alternative data sources including proxy labour measures from administrative data.

Contact:
Chris Daffin
chris.daffin@ons.gov.uk
+44 (0) 1633455858

Release date:
20 September 2018

Next release:
To be announced

Table of contents

1. [Disclaimer](#)
2. [Executive summary](#)
3. [Overview](#)
4. [Aggregate-level analysis](#)
5. [Unit level analysis](#)
6. [Conclusions and recommendations](#)
7. [References](#)
8. [Appendix: Data sources and study variables](#)

1 . Disclaimer

These findings are NOT official statistics on income and benefits. Rather they are published merely as findings from research into a different methodology to that currently used in the production of the Labour Force Survey (LFS) estimates.

It is important the information and research presented here is read alongside the outputs to aid interpretation and avoid misunderstanding. These outputs must not be reproduced without this disclaimer and warning note.

2 . Executive summary

The Labour Force Survey (LFS) is experiencing an ongoing decline in response rate, from over 80% in the early 1990s to approximately 55% in 2017. As the response rate declines the risk of non-response bias increases. In 2012, the Census Non-Response Link Study (CNRLS) estimated that LFS bias produced an under-estimation of employment by approximately 110,000 (where the total employment was 26 million). Given that this was a relatively small under-estimation, accompanied by the fact that response patterns evolve over time, it was recommended not to introduce a non-response weight based on the 2011 Census.

In 2017, Office for National Statistics (ONS) initiated a new project to explore administrative and wider data as a means of assessing the potential bias in the LFS more regularly than once every 10 years, when census data is available. The project has two parts:

- firstly, a comparison of the LFS – against primarily published aggregate estimates – on broader employment-related aspects
- secondly, a feasibility study to assess the extent to which administrative datasets could provide proxy employment and unemployment estimates

So far, the study has compared LFS measures of employment and unemployment to proxies derived from annual Department for Work and Pensions (DWP) administrative data. The study shows that the DWP data in isolation could not be a direct replacement for the LFS measures. Further investigation would be needed using point in time rather than annual data from DWP to ascertain if this would be a better match. Also, further work could be completed to align concepts and definitions between the ONS outputs and the administrative data systems.

The first part of the project, that is, comparison of LFS estimates with those from other data sources on aspects such as Workforce Jobs, employment by public sector and Jobcentre Plus, shows a consistent pattern of little or no increased deviation over time.

However, the second part of the study highlighted the potential for using DWP data for comparing LFS responders versus non-responders, through linking datasets together. It found that households present on Pay As You Earn (PAYE) had a notably higher response rate on the LFS. This analysis should be extended as it conflicts with the CNRLS analysis, which showed a similar bias but in the opposite direction.

Despite the difference in profile shown in the Labour Force Survey-Benefits and Income Data (LFS-BIDS) linked dataset, the impact on LFS estimates of employed people is calculated as an over-estimate of 93,000 for England and Wales. This compares with a CNRLS calculation of a 110,000 under-estimation of employment. In other words, while the direction of bias in the weighted sample may have changed, it is broadly unchanged in its magnitude.

More data, such as Pay As You Earn-Real Time Information (PAYE-RTI) and Self-Assessment data from Her Majesty's Revenue and Customs (HMRC), is needed to investigate this further.

The linking part of the project has provided useful information on the difference between responders and non-responders to the LFS, using data from England, Scotland and Wales. Combining this analysis with the analysis of other data sources, we conclude that so far, the study has found no clear evidence of increased bias in the achieved LFS sample since 2011. However, further exploration of wider data needs to be conducted to confirm or reject this and gain a better feel for its potential benefit as a source of assessing and addressing bias.

We recommend the following actions in the future:

1. As a result of this analysis, a non-response bias adjustment should not be made to the LFS.
2. Arrange access to PAYE and DWP data more frequently to have data for points in time throughout the year rather than annual extracts.
3. Investigate the differential match rate between responding and non-responding households using DWP extracts from time-points close to the LFS interview dates.
4. Repeat the bias analysis by linking LFS samples to HMRC PAYE-RTI data.
5. Investigate the robustness of DWP data-based non-response adjustment factors by repeating the analysis for several quarters.
6. Explore wider administrative, big and commercial data sources as a potential basis for identifying bias in the achieved LFS sample.
7. The LFS team should work closely with the Social Statistics Transformation team to ensure that ongoing analysis feeds both the ongoing monitoring of LFS outputs and the work to design the new “administrative data first” social statistics system.

This research should be ongoing, with new data sources being identified and accessed as part of ONS's focus on transforming statistics through new and improved methods.

3 . Overview

Background and research objectives

It is well understood that long-term trends in response rates for Office for National Statistics (ONS) social surveys are downwards; declining from around 80% in the early 1990s to below 60% today. This is broadly consistent with evidence from other countries and the conclusion that response rates were declining in North America and Europe by approximately one percentage point per year.

The reasons for this trend fall broadly into two categories: those which explain increases in the refusal rate and those explaining the rising non-contact rate. For ONS surveys the refusal rate has increased more than the non-contact rate. However, response rates do differ across surveys, which can be largely explained by different survey characteristics; for example, the length of field period, questionnaire length, single or multiple respondents per household, expenditure diary and so on.

Other factors that affect response that are common to all surveys include the following:

- falling contact rates, which may be attributed to changing socio-demographics, a rise in controlled access to properties, or interviewer behaviours, including loss of expertise through high interviewer turnover
- survey and data overload: the salience of a survey topic has become a more important determinant of response
- external factors such as attitudes to government or information losses
- pressure on budgets, which result in less appetite to administer costly re-issue exercises or introduce or increase monetary incentives

The ONS Labour Force Survey (LFS) is experiencing a similar decline in response rate, from over 80% for wave 1 cases in 2001 to approximately 55% in 2017. As the response rate declines, there is an increased risk of non-response bias occurring, where the potential differences between the characteristics of responders and non-responders lead to estimates being distorted, so that they no longer accurately reflect the population. It is important that we understand non-response and do everything we can to reduce its potential effects on estimates.

When subpopulation groups with different profiles with respect to the survey variables of interest, for example, high employment in one group and low employment in another, show different response rates, this can result in bias in the survey estimates. The size of the bias can be important when the differences in profile and response rates are large and the subpopulation groups aren't too small.

In 2012, the 2011 Census Non-Response Link Study (CNRLS) (see Weeks and others, 2013) considered whether there was evidence of bias in the achieved LFS sample at that time. It identified some concern over the incidence of some sociodemographic groups in the achieved wave 1 sample and estimated an under-estimation of employment by approximately 110,000 (0.3%) in the LFS. However, the relatively small impact on employment estimates combined with evolving response patterns led to a recommendation not to introduce a non-response weight based on factors derived from census findings. Instead it was recommended to explore the potential use of administrative data sources in the weighting procedure to adjust for differential non-response.

In 2017, a new project was initiated to review administrative, and other published, data to investigate the feasibility of assessing (and potentially addressing) non-response bias in the LFS. This is particularly pertinent given that the LFS wave 1 response rate has fallen from 62% in 2011 to approximately 55% in 2017. This report provides initial findings, focusing on the feasibility of using wider data to assess non-response bias in the achieved LFS sample.

Approaches adopted

Two approaches have been adopted in this work, one at aggregate level and another at unit level. In the aggregate level approach, aggregate estimates produced from the LFS microdata have been compared with aggregate estimates from external sources, focusing on more detailed characteristics such as the number of jobs and the sector of employment.

In the unit level approach, the LFS sample for England, Scotland and Wales has been linked to Department for Work and Pensions (DWP) Customer Information System (CIS) and Benefits and Income Data (BIDS), with a focus on defining proxy measures in those administrative datasets to compare responding and non-responding addresses and derive non-response adjustment factors.

In addition, given that use of administrative data for this type of analysis is in its infancy, the report also considers how this type of analysis might be extended to wider administrative and big data sources.

4 . Aggregate-level analysis

We consider two other sources for labour market statistics; a source based mostly on a business survey, the Workforce Jobs series and two sources based on an administrative source, Jobcentre Plus and the NHS Workforce.

Workforce Jobs series

As part of the quarterly release of labour market statistics Office for National Statistics (ONS) publishes a comparison of Labour Force Survey (LFS) estimates with estimates from the Workforce Jobs series, which is based mostly on the Monthly Business Survey. See the National Statistics Quality Review Series Report Number 44 for details of the methodology used to compare the two data sources.

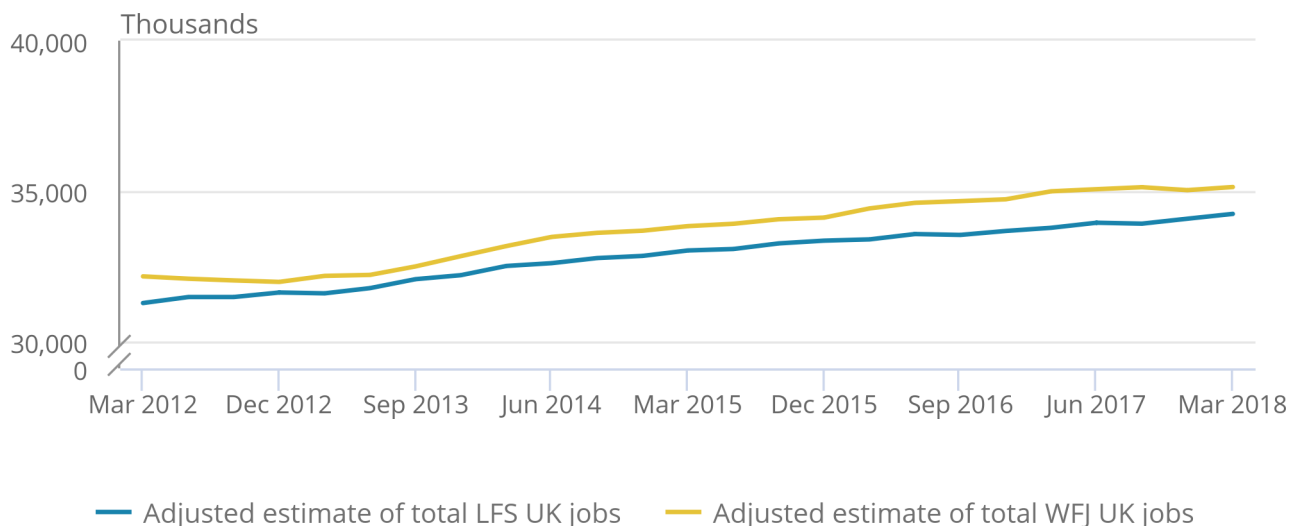
As can be seen from Figure 1, the LFS series is consistently below the Workforce Jobs series (note that the two series measure slightly different parameters), but the gap between the two series hasn't increased notably over time.

Figure 1: Labour Force Survey and Workforce Jobs estimates of jobs adjusted for measurable differences, thousands (seasonally adjusted)

UK, March 2012 to March 2018

Figure 1: Labour Force Survey and Workforce Jobs estimates of jobs adjusted for measurable differences, thousands (seasonally adjusted)

UK, March 2012 to March 2018



Source: Labour Force Survey and Workforce Jobs, Office for National Statistics

Notes:

1. Comparison of number of jobs from the Labour Force and Workforce Jobs Surveys

It is noted that both Workforce Jobs and the Inter-Departmental Business Register provide details of number of jobs or employees by sector. This is useful information to compare against LFS estimates by sector and should be developed further.

Admin sources: Jobcentre Plus and NHS Workforce

Jobcentre Plus series

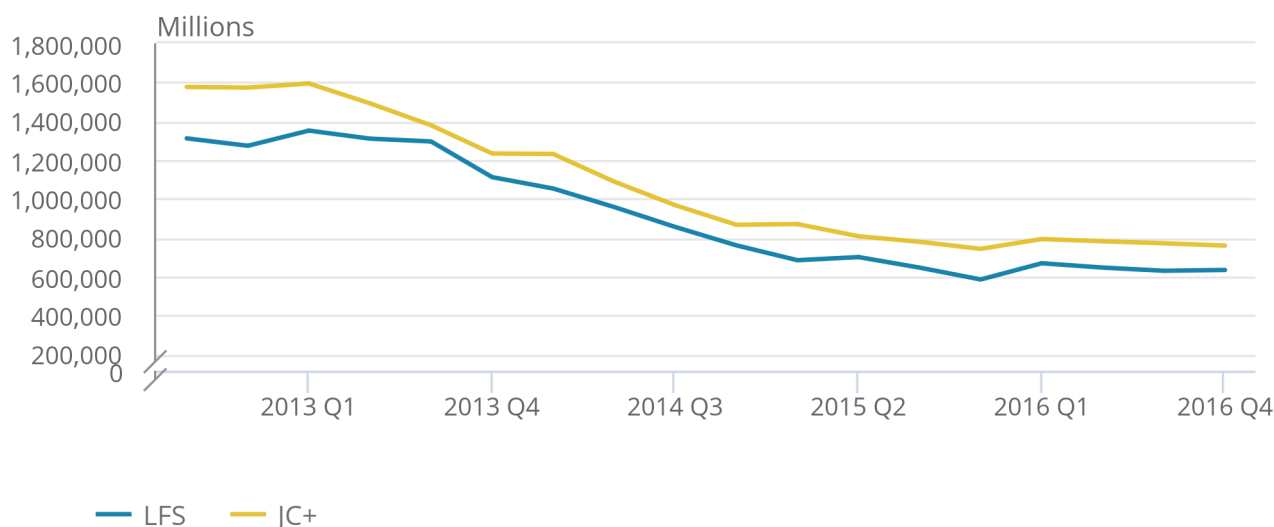
The Jobcentre Plus series is published by Department for Work and Pensions (DWP); it measures the number of people in receipt of Jobseeker's Allowance (JSA). From the LFS, estimates of people on JSA, as reported by responders, can also be calculated. Figure 2 shows the DWP and LFS series between Quarter 3 (July to Sept) of 2012 and Quarter 4 (Oct to Dec) of 2016; the gap between the two series doesn't show an obvious systematic increase, which indicates that the LFS suffers from little bias with respect to JSA estimates.

Figure 2: Time series of Jobseeker's Allowance statistics from Labour Force Survey and Jobcentre Plus

UK, Quarter 3 (July to Sept) 2012 to Quarter 4 (Oct to Dec) 2016

Figure 2: Time series of Jobseeker's Allowance statistics from Labour Force Survey and Jobcentre Plus

UK, Quarter 3 (July to Sept) 2012 to Quarter 4 (Oct to Dec) 2016



Source: Labour Force Survey, Office for National Statistics and Jobcentre Plus, Department for Work and Pensions

Notes:

1. The gap between the Labour Force Survey and Job Centre Plus shows no obvious systematic increase, indicating that the LFS has little bias with respect to JSA estimates

NHS Workforce

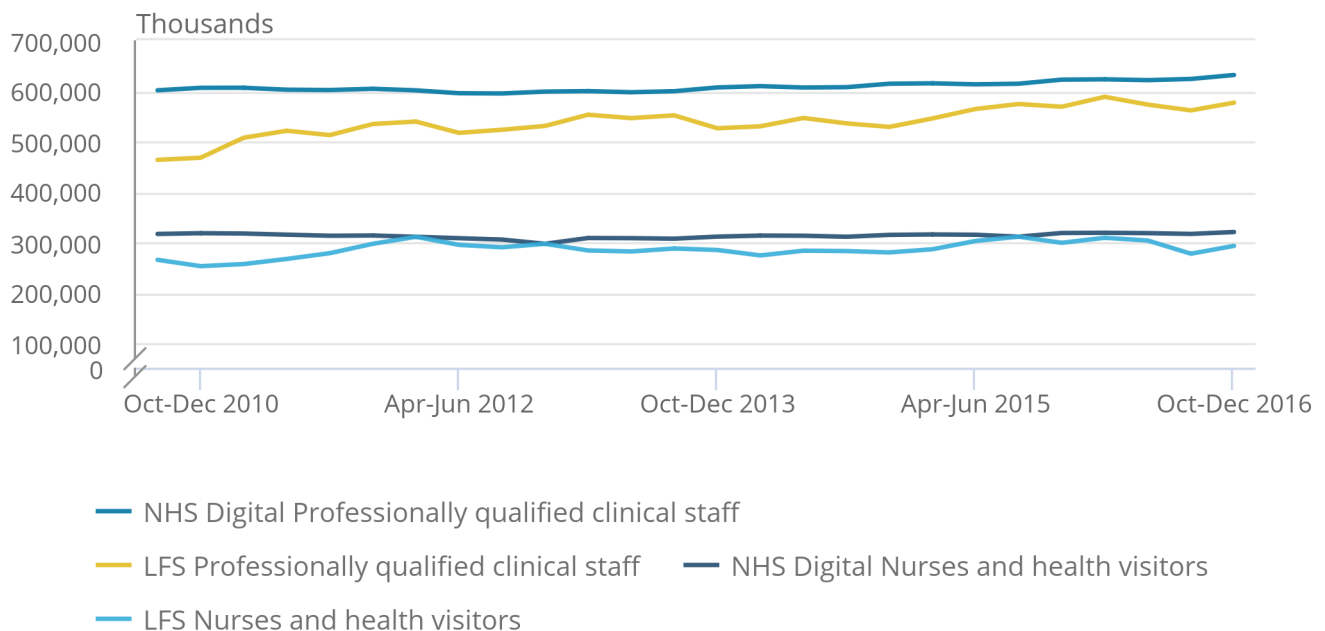
NHS Digital publishes NHS Workforce Statistics. Figure 3 shows a comparison between headcount figures and LFS estimates for professionally-qualified clinical staff, and nurses and health visitors, for calendar quarters between July to September 2010 and October to December 2016.

Figure 3: Comparison between NHS Workforce Statistics and Labour Force Survey estimates for professionally-qualified clinical staff, and nurses and health visitors

England, July to Sept 2010 to Oct to Dec 2016

Figure 3: Comparison between NHS Workforce Statistics and Labour Force Survey estimates for professionally-qualified clinical staff, and nurses and health visitors

England, July to Sept 2010 to Oct to Dec 2016



Source: Labour Force Survey, Office for National Statistics and NHS Workforce Statistics, NHS Digital

Notes:

1. The changing nature of organisations that provide NHS services as part of Transforming Community Services (TCS) may impact on the overall workforce totals as a greater number of third party providers external to the NHS are excluded from the figures.
2. On the LFS, workers and agency staff who provide mainly NHS services will be included in NHS employment estimates.

This indicates a small closing of the gap between LFS and NHS Digital estimates in relation to professionally-qualified clinical staff. The nature of this pattern, and possible reasons for it, might be considered further.

5 . Unit level analysis

In this section, we start by describing the linking of the Labour Force Survey (LFS) sample to Benefits and Income Data (BIDS), then describe the methodology of bias assessment and finish by presenting the results of the analysis on the assessment of bias.

Linking LFS samples with BIDS

The eligible addresses of the first wave in each of three annual LFS samples, for the tax years (April to March) ending 2014, 2015 and 2016, for both respondents and non-respondents, were linked to extracts of the Customer Information System (CIS) database taken in June of 2013, 2014 and 2015, respectively. Each year had around 60,000 addresses of which about 55% were responders. The link rates between the LFS-sampled addresses and CIS are given in Table 1. We can see that the link rates of responding addresses are slightly higher than those of non-responding addresses, which may impact the results of this analysis.

Table 1: Link rates between Labour Force Survey samples and Customer Information System, UK, tax year ending 2014 to tax year ending 2016

Year	Responding addresses	Non-responding addresses
2013 to 2014	90.2	88.2
2014 to 2015	90.9	89.2
2015 to 2016	91.3	88.9

Source: Labour Force Survey, Office for National Statistics and Customer Information System, Department for Work and Pensions

The linked addresses were then linked to the BIDS database, which includes Pay As You Earn (PAYE) income from HM Revenue and Customs (HMRC) for each financial year. The linking was done at address level as we needed to analyse the profiles of both responding and non-responding addresses, and for the latter no names were available to undertake the linking. For the purposes of bias assessment, the results are based on the household reference person (HRP) only. This was done this way as we needed to apply factors at household level and the status of the HRP is a good predictor of household response and easy to obtain.

At each linked address an HRP was defined in BIDS: it is the person with the highest PAYE income or the oldest person, if nobody at the address has a PAYE income. The LFS-BIDS linked sample was partitioned into two groups: a group where the HRPs had a positive PAYE income and another group where the HRPs had zero PAYE income or were not present on the PAYE system.

Table 2 gives the response rates for each of the two groups (the response rates were computed on the set of linked addresses). Note that the group “Zero PAYE income/not on PAYE system” is dominated by addresses where the HRP is not on the PAYE system (about 85% of the addresses).

Table 2: Response rates in relation to PAYE income of household reference persons in Labour Force Survey-Benefits and Income Data linked datasets, UK, tax year ending 2014 to tax year ending 2016

Response rates (percentage)			
Year	PAYE income	Zero PAYE income/ not on PAYE system	Difference
2013 to 2014	58.8	52.8	6.0
2014 to 2015	59.4	52.6	6.9
2015 to 2016	55.4	50.4	5.0

Source: Labour Force Survey, Office for National Statistics and Benefits and Income Data, Department for Work and Pensions

The response profile by the PAYE status can be compared with that in the Census Non-Response Link Study (CNRLS) in 2001 and 2011. The CNRLS looked at characteristics of responders and non-responders in LFS in relation to socio-economic characteristics collected by the census, including employment. Table 3 shows that in the 2011 CNRLS, households where the HRP was employed were less likely to respond than households where the HRP was not employed, by nearly 2 percentage points. In the 2001 CNRLS, the difference was in the same direction as in the current LFS-BIDS study but it is much smaller in magnitude.

Table 3: Response rates in the 2011 Census Non-Response Link Study, UK, 2001 and 2011

Year	Employed	Not employed	Difference
2011	61.30%	63.10%	-1.80%
2001	80.60%	80.40%	0.20%

Source: Labour Force Survey, Office for National Statistics

Taking inclusion in the PAYE system as a proxy for being in employment, the non-response gap is found to be much larger in tax year ending 2014 than in the 2011 CNRLS. Given that less than three years separate the date of the 2011 Census from the first year of the BIDS data (tax year ending 2014), the difference is unlikely to be the result of an increase in bias; it is probably caused by differences of coverage between the two data sources (the LFS sample selected to represent a very specific target population) and the use of different proxies for employment, as well as the limitations of the linking process given the data available (see following section).

Derivation and application of non-response adjustment factors

Non-response factors were computed for the following age groups of the HRP: 16 to 24, 25 to 44 and 45 to 64 years. For the PAYE-based non-response factors to have an impact on non-response bias assessment, there should be a strong association between PAYE income and employment status. However, because we could not distinguish PAYE income from retirement and employment, and a larger proportion of those who are 65 years or older are in retirement, we decided that no factors should be calculated for this age group.

A non-response factor in an age group is defined as the inverse of the response rate in that age group. The non-response factors were scaled so that their average is equal to 1.

Table 4 gives the non-response factors for three years. The PAYE indicator is set to 1 for addresses where the HRP has a positive PAYE income value; it is set to 0 for addresses where the HRP has either zero PAYE income value or is not on the PAYE system. The non-response factors are mostly below 1 for households where the HRP received PAYE income, which indicates oversampling, and above 1 for households where the HRP did not receive PAYE income, which indicates under-representation. The factors in the latter group do show variation from year to year.

Table 4: Labour Force Survey non-response factors based on Benefits and Income Data, UK, tax year ending 2014 to tax year ending 2016

Age group	PAYE Indicator	2013 to 2014	2014 to 2015	2015 to 2016
16 to 24	1	1.00	1.01	0.99
25 to 44	1	0.94	0.95	0.94
45 to 64	1	0.91	0.91	0.92
16 to 24	0	1.07	1.01	1.11
25 to 44	0	1.06	1.06	1.02
45 to 64	0	1.01	1.06	1.02

Source: Labour Force Survey, Office for National Statistics and Benefits and Income Data, Department for Work and Pensions

Applying non-response factors to the tax year ending 2016 LFS dataset

The linked addresses of the wave 1 LFS samples were assigned the PAYE income value of the HRP, or a missing value if the defined HRP is on the PAYE system. An indicator showing whether the HRP has a positive income value or not was added to the tax year ending 2016 dataset, which was used to estimate the bias in main labour market estimates. Depending on the PAYE indicator of an address, a non-response factor was assigned to each address. Addresses that were not linked to the CIS or where the HRP is 65 years old or over were assigned a non-response factor equal to 1. As a result, about 18% of adults under the age of 65 years had no non-response adjustment applied to their data.

The LFS is a self-weighted sample within Great Britain and Northern Ireland, that is, all households have an equal design weight within Great Britain and Northern Ireland, which can be calculated simply as the size of the Great Britain and Northern Ireland population by the respective sizes of the responding samples. The design weights were then multiplied by the non-response factors to obtain pre-calibration weights. Calibration to population totals was then performed using the usual partitions of the LFS.

A separate weighting of the wave 1 dataset was performed without applying the non-response adjustment factors.

Estimates of economic activity outcomes (International Labour Organization definition) were obtained with and without a non-response factor; the difference between the latter and the former estimates gives an estimate of the bias. The difference between the estimates without and with a non-response adjustment gives an estimate of the bias. Table 5 shows the estimates of bias for economic activity in the tax year ending 2016. We can see there was an overestimate of employment of about 103,500 in the tax year ending 2016, which represents about 0.3% of total employment and an underestimate of unemployment of about 13,000, which represents about 0.8% of total unemployment.

Table 5: Estimates of bias in International Labour Organization outcomes of UK economic activity estimates in tax year ending 2016

Economic Activity	Estimate without a non-response adjustment	Estimate with a non-response adjustment	Bias estimate	Percentage bias
Employed	31,095,456	30,991,962	103,494	0.33
Unemployed	1,674,737	1,687,962	-13,225	-0.78
Inactive	19,386,126	19,476,106	-89,980	-0.46

Source: International Labour Organization

Note that the 2011 CNRLS showed an underestimate of employment of about 110,000 for England and Wales. The equivalent figure for the tax year ending 2016 based on LFS-BIDS linking is an overestimate of about 93,000. For unemployment, it showed an overestimate of about 1,000.

Limitations of unit-level analysis

Income data from the PAYE system were available to ONS. This was restricted to an individual's total gross income from employment and pension per tax-year. There are components of income that are missing, for example, income from self-employment and investments taxed via self-assessment. This means that the PAYE indicator used in this analysis is less well associated with LFS ILO employment as the employment status collected in the census, which was used in the CNRLS. Also, separating PAYE income from employment and a pension should make the indicator more effective at bias assessment.

Sources of linkage error

The linking is subject to two main errors.

Firstly, LFS to CIS linkage was performed using the unique property reference number (UPRN). This is derived using a georeferencing code from Address Base.

There were a small proportion of cases (approximately 2%) where the georeferencing process failed, that is, the LFS addresses did not resolve to a UPRN. In some circumstances, record linkage may have occurred at a higher level, that is, at building level rather than an individual property. This may introduce an additional level of error in the analysis as individuals may not in fact reside at a household but instead reside in a neighbouring property in the same address level.

Also, an address record in CIS can contain several households¹ but on LFS only one household is selected in a multi-household address. This particularly relates to student housing and communal establishments.

Additionally, where a UPRN could not be assigned to an address, a parent UPRN (which relates to the entire floor or building) was assigned. Both produced some linked addresses with very high occupancy. To reduce this error, addresses on CIS that contain more than 12 persons were excluded from the analysis.

Secondly, another source of error relates to coverage; the linkage rate between LFS and CIS is slightly higher in the group of responding addresses of the sample (see Table 1). This could be explained by the fact that the self-employed, who may be less likely to be on CIS than individuals with another status, are over-represented in the group of non-responders.

Finally, a linked address may not contain the same household that was sampled in LFS because of timing (the CIS extract relates to a specific day of the year whereas the LFS sampled addresses are contacted throughout the year. For instance, we found that about 16% of the responding households where the HRP is aged 65 years or older were linked to households in BIDS where the HRP is less than 65 years old.

It is unclear to what extent this type of error has affected the non-response factors we calculated and the impact on bias; the analysis based on real time information (RTI) data should address this problem.

Notes for: Unit level analysis

1. A household comprises a single person, or a group of people living at the same address who have the address as their only or main home. They also share one main meal a day or share the living accommodation (or both).

6 . Conclusions and recommendations

Summary of findings

The Jobseeker's Allowance (JSA) data from Jobcentre Plus proved a usable series for comparison of aggregate level data and it displays a similar trend over time to the Labour Force Survey (LFS).

Linking of LFS and Benefits and Income Data (BIDS) data at address level produced a fairly high match rate overall and offers potential. The addresses at which an LFS response was obtained have a higher match rate than the non-responding addresses. Although we expect a lower match rate in the non-responding category, as it includes ineligible addresses, it is unclear whether the ineligibility rate is consistent with the difference in match rates. Further data is needed to investigate this.

Comparisons of the responding and non-responding addresses with respect to Pay As You Earn (PAYE) show:

- the response rate of the addresses with PAYE income is higher by about 6 percentage points than the response rate of addresses that have no PAYE income – this is seen in all age groups; this translates into a bias estimate of about 103,500 in the tax year ending 2016
- the difference in response rates between the "PAYE" and "No PAYE" addresses does not appear to have widened between the tax year ending 2014 and the tax year ending 2016
- this contrasts with a similar analysis conducted on an LFS sample that was linked to 2011 Census data; it showed very little difference in response rates with respect to employment (as recorded in the census); this flags that further analysis is required to understand the value of the of BIDS data for use in relation to the LFS, which focuses on point-in-time estimates

Aggregate level analysis highlights some differences between estimates derived from the LFS versus those from other published sources. However, there is little sign of the disparity between estimates widening:

- Workforce Jobs produces an estimate of approximately 3% more jobs (when measurable factors causing difference are removed) than the LFS measure of main and second job combined; the gap between the two series appears to be slightly greater in recent periods, but this may be due to changing employment patterns
- when assessment is made of employment profiles in LFS versus other government department publications we see a close alignment for each of police workforce, secondary teachers, and nurses and health visitors
- there is less alignment for nursery and primary school teachers, and professionally-qualified clinical staff; however, the gap is not increasing in either case
- there is also strong ongoing alignment of mean earnings estimates between LFS and average weekly earnings (AWE)

Conclusions

The comparison of LFS estimates for the employment and unemployment proxies JSA and Employment and Support Allowance (ESA) with the aggregate statistics for these variables from BIDS does not allow us to assess the potential bias in the LFS. The current PAYE annual construction is not suitable for providing such proxy indicators required for application to point-in-time estimates, which are the focus of the LFS.

However, the high match rate of LFS and BIDS cases indicates potential use of the data for comparing the LFS responding and non-responding household sets. The analysis of the tax year ending 2016 BIDS, which suggests that the LFS sample suffers from some bias with respect to employment and unemployment proxies in BIDS, in particular PAYE, is notable. However, the fact that the most recent analysis of LFS-census linked data showed a much less biased sample indicates that the amount of bias in the present analysis has been exacerbated by the differences in coverage between BIDS and the Postcode Address File (PAF) and linkage error. More data, such as RTI data, should be utilised to investigate this further.

Despite the suggested bias of the LFS sample with respect to the BIDS proxies, this does not appear to translate to notable bias in the weighted LFS measures. Further, comparison of LFS estimates with other published series shows no clear evidence of a divergence between estimates, that is, no sign of increased bias in the LFS.

This work represents initial steps in assessing the use of administrative and wider data for identifying and addressing bias. There are numerous other data sources that will become available and should be utilised.

Recommendations and next steps

1. As a result of this analysis a non-response bias adjustment should not be made to the LFS.
2. Arrange access to PAYE and Department for Work and Pensions (DWP) data more frequently to have data for points in time throughout the year rather than annual extracts.
3. Investigate the differential match rate between responding and non-responding households using DWP extracts from time-points close to the LFS interview dates.
4. Repeat the bias analysis by linking LFS samples to HM Revenue and Customs (HMRC) Pay As You Earn-Real Time Information (PAYE-RTI) data.
5. Investigate the robustness of DWP data-based non-response adjustment factors by repeating the analysis for several quarters.
6. Explore wider administrative, big and commercial data sources as a potential basis for identifying bias in the achieved LFS sample.
7. The LFS team should work closely with the Social Statistics Transformation team to ensure that ongoing analysis feeds both the ongoing monitoring of LFS outputs and the work to design the new “administrative data first” social statistics system.

This research should be ongoing, with new data sources being identified and accessed as part of ONS’s focus on transforming statistics through new and improved methods.

7 . References

Amber Weeks, Andy Fallows, Peter Broad, Salah Merad and Karl Ashworth (2013) [Non-response Weights for the UK Labour Force Survey? Results from the Census Non-Response Link Study \(PDF, 169KB\)](#).

[National Statistics Quality Review Series. Report No. 44 \(PDF, 4.4MB\)](#)

8 . Appendix: Data sources and study variables

The data sources used in the analysis comprise:

1. Labour Force Survey (LFS) responses for the tax years (April to March) ending 2014, 2015 and 2016.
2. LFS-sampled addresses for the same years, where each address has an outcome code, which indicates whether the household co-operated with the survey or not.
3. Customer Information System (CIS) extracts for 2013, 2014 and 2015, where the extract dates are 30 June 2013, 30 June 2014 and 3 July 2015, respectively.
4. Benefits and Income data (BIDS)* for the tax years ending 2014, 2015 and 2016.
5. Jobcentre Plus administrative data, which have been published as a time-series by Office for National Statistics (ONS) (these are classified as experimental statistics).

*Customer Information System (CIS) is a cumulative system of all individuals who have registered and been issued with a National Insurance number, of which ONS receive annual extracts.

**Benefit and Income Data (BIDs) is a set of datasets sourced from both the Department for Work and Pensions (DWP) and Her Majesty's Revenue and Customs (HMRC). The full list consists of:

- Single Household Benefit Extract (SHBE) – which contains data from local authority housing; this data relates to Housing Benefit claimants (and any partner dependants), payments and awards
- National Benefits Database (NBD) – which holds data on 13 different DWP and HMRC benefits (excluding Universal Credit); in particular start and end dates of the benefits claimed are provided allowing the “activity” of individuals to be built up, this activity data allows the possibility to improve the quality of address information by mapping to the latest address
- HMRC Pay As You Earn (PAYE) data on individuals for a given tax year; this is a derived dataset holding all the PAYE pay from all employments and pensions
- HMRC New Tax Credits (TC) – data on payments to individuals including Child and Working Tax Credits with a tax year
- HMRC Child Benefit (CB) – derived to make the dataset “child-based” with the main payee or customer linked to each child; this will cater for any number of children within a claim, the data is supplied by DWP with the associated start and end date
- Universal Credit (UC) flag data with associated start and end date of claims
- Personal Independence Payment (PIP) flag data including associated start and end dates of claims

The Jobcentre Plus series is available as both seasonally adjusted and not seasonally adjusted and includes estimates of Jobseeker's Allowance (JSA) claimants and Universal Credit (UC) claimants. Estimates are given for each month ranging from January 2013 to May 2017. We used the non-seasonally adjusted time series.

The LFS-BIDS analysis focused on employment proxy variables in BIDS, including those relating to benefits, such as Pay As You Earn (PAYE) income. We attempted to assess the potential bias in the main LFS measures of employment and unemployment.