Office for
Statistics Regulation

UK Statistics
Authority

Assessment of compliance with the Code of Practice for Statistics

Statistics on earnings and employment from Pay As You Earn Real Time Information (PAYE RTI), UK

(produced by HM Revenue and Customs and Office for National Statistics)

Assessment Report 377

June 2023

Office for Statistics Regulation

We provide independent regulation of all official statistics produced in the UK. Statistics are an essential public asset. We aim to enhance public confidence in the trustworthiness, quality and value of statistics produced by government.

We do this by setting the standards they must meet in the [Code of Practice for Statistics.](#) We ensure that producers of government statistics uphold these standards by conducting assessments against the Code. Those which meet the standards are given National Statistics status, indicating that they meet the highest standards of trustworthiness, quality, and value. We also report publicly on system-wide issues and on the way statistics are being used, celebrating when the standards are upheld and challenging publicly when they are not.

Executive Summary

ES.1 The Office for Statistics Regulation (OSR) started a regulatory review of HM Revenue and Customs (HMRC) and Office for National Statistics (ONS) [estimates of earnings and employment from Pay as You Earn Real Time Information (PAYE RTI) UK](#) (to be referenced as PAYE RTI) in October 2022, which are currently labelled as 'experimental statistics'. The aim of our review was to provide feedback and support HMRC and ONS's endeavours to further develop these statistics, with a view to achieving National Statistics status in the future. To maximise the value of OSR findings and in response to advice, HMRC and ONS requested an Assessment against the [Code of Practice for Statistics](#) to set out a clear path to achieving [National Statistics](#) status.

ES.2 We have identified six requirements for HMRC and ONS to fulfil for the earnings and employment statistics from PAYE RTI to be designated as National Statistics. Once HMRC and ONS demonstrate that these steps have been taken, OSR will recommend that the UK Statistics Authority designate the statistics as National Statistics.

ES.3 We consider it would help further promote confidence in these statistics for HMRC and ONS to publish a plan that sets out their proposals for addressing the requirements.

What we found

ES.4 The development of the monthly estimates of earnings and employment using PAYE RTI data has proved valuable to users, filling a timely data gap especially during the COVID-19 pandemic and providing benefits over existing labour market data.

ES.5 There is huge potential to gain further value from this rich data source through data linkage and re-use opportunities which should be exploited by both teams. The ability to match these data with other sources, as demonstrated by the existing linkage to the ONS's Inter-Departmental Business Register, also offers the potential to produce better statistics more cheaply.

ES.6 There are many sources of labour market data available, but it is not clear how coherent the PAYE RTI estimates are with these other sources. ONS should look to provide a full picture of the labour market data sources for users to be able to determine which sources are best suited for their need.

ES.7 Engagement around the PAYE RTI estimates within government is good with key users able to hear about developments and feed in their views. We encourage HMRC and ONS to look to engage with a broader range of labour market statistics users on an ongoing basis, to help foster a wider user dialogue.

ES.8 While information about the methodology is published, including recently made changes to the imputation methods, we found that this material could be improved, and this was reflected in the mixed views we received from users on its value. HMRC and ONS should consider different user needs when presenting methodological information making sure that it is relevant and targeted at the right level. HMRC also needs to do more to reassure users about its quality assurance.
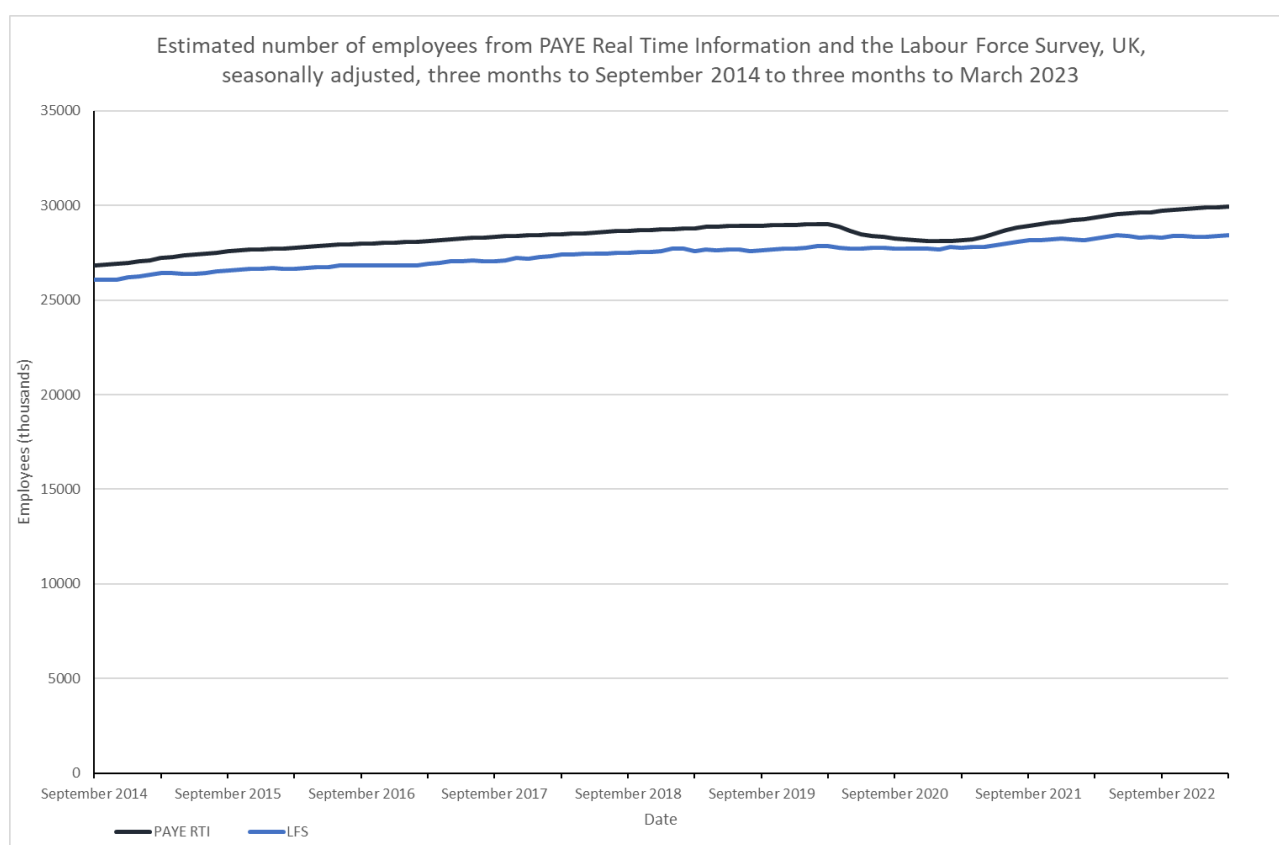
Findings

Introduction

1.1 Data for employment and earnings estimates are collected via HMRC's Pay As You Earn Real Time Information (PAYE RTI) system. This is the source for the [UK estimates of earnings and employment from PAYE RTI](), which are produced jointly by the HM Revenue and Customs (HMRC) and the Office for National Statistics (ONS) and are published monthly by ONS at the same time as its other [Labour Market statistics.]() These estimates were developed from HMRC's previously published quarterly estimates using PAYE RTI data which were then paused to allow for development work and collaboration with ONS to improve the statistics and better meet user needs. HMRC owns the PAYE RTI data and is responsible for the collection, processing, and quality assurance of the data. HMRC sends the finalised aggregate data tables to ONS for it to publish on its website.

1.2 These statistics provide a timely indicator of trends in earnings and employment in the UK labour market. They were particularly valuable early in the COVID-19 pandemic as they provided much more responsive insights into the UK labour market than existing statistics. They are regularly used within the UK Government and the devolved administrations. They also provide valuable insights into geographical inequalities in earnings and employment and allow comparisons across subnational labour markets, as well as being able to look at the different regional patterns in the impact of, and recovery from, the COVID-19 pandemic. These statistics are also used by other organisations such as charities and think tanks who have a specific interest in the labour market allowing them to carry out their own analyses on topics such as income and its impact on living standards and the growth in pay.

1.3 The PAYE RTI estimates form part of the wider suite of labour market statistics that are also used to inform a range of government policies and decisions, for example to help target particular groups such as older workers and jobless households participate in the labour market and find work. The use of an administrative source enables the statistics to have wider coverage and be more complete, whereas a survey would provide data for only a sample of people or businesses.

1.4 At the start of the COVID-19 pandemic in April 2020, and in response to user need for more timely data on employees, HMRC introduced an early estimate of the number of employees paid through the tax system for the previous month, with a lag of less than three weeks, sometimes referred to as the 'flash estimate'. The flash estimate is largely built on received employee records, around 85%, with the residual records imputed based on historic patterns. The imputations decrease to less than 1% of records in the following months.

1.5 Another benefit of the statistics is that they provide a timelier breakdown compared with other ONS labour market data sources such as the Annual Survey of Hours and Earnings (ASHE), as well as more-granular data compared with the Average Weekly Earnings data (AWE), Labour Force Survey data (LFS) and Workforce Jobs data (WFJ) – with a monthly publication cycle and a time lag of around six to seven weeks after the end of the reference period.

Understanding the landscape

1.6   Users we spoke to as part of our review recognise and emphasised the value that these statistics offer. While some users we spoke to find these timelier estimates beneficial, others expressed valid concerns about the systemic downward revisions that were occurring each month in the early estimates. We welcome that HMRC has updated its imputation methodology which is reducing the magnitude of the revisions (see para 1.18 for further details).

1.7   As outlined in the ONS income and earnings statistics guide, there are many sources of labour market data available, including the PAYE RTI statistics. Each data source has its own purpose, definition and meaning, and this crowded landscape can make it difficult for users to navigate and identify what the most appropriate figure is for their individual needs. For example, estimates of employees using PAYE RTI data are consistently higher than estimates of employees from the Labour Force Survey (LFS) (see chart below) due to the different data collection methods and definitions of employees.

Estimated number of employees from PAYE Real Time Information and the Labour Force Survey, UK, seasonally adjusted, three months to September 2014 to three months to March 2023

Source: Office for National Statistics – Labour Force Survey; HM Revenue and Customs – Pay As You Earn Real Time Information. Note a three-month rolling average of RTI payrolled employees has been constructed for this comparison.

1.8   It is good to see that an article published on the ONS website in July 2022 includes visual comparisons and explanations of the differences with other labour market statistics, such as employee numbers and employee growth between the PAYE RTI and LFS. However, the data in this article has not been updated since 2019. There is also a separate more-thorough Comparison of Labour Market data sources. We agree with those expert users who are familiar with the data and the landscape, who expressed a concern about the lack of an explanation into the coherence and

comparability of the labour market data sources, particularly for those who are less familiar with the data, and that clear guidance on which figure should be used for what circumstances would be helpful. A good example of providing such information on the data landscape and uses is DLUHC's guide to its housing statistics.

Requirement 1: To help users navigate and understand the coherence and comparability of labour market data available ONS should explain how the PAYE RTI statistics relate to other equivalent labour market statistics in a way that helps the full range of users understand how the different data sources relate to each output and the respective strengths of different figures to answer key questions about the UK's labour market.

User engagement

1.9   HMRC hosts a quarterly RTI steering group which offers a good channel to engage with key users across UK Government and the devolved administrations and helps to maintain a consistent dialogue with known users. The group is used to inform users of any recent changes (for example methodological changes), plans and for discussion of any issues to be raised. While some users that we spoke to who attend the group found it useful and provided positive feedback, some users would welcome additional engagement activities to support their ongoing data queries. Some users within the group were unclear on responsibilities across HMRC and ONS and were unsure where queries about different aspects of the data should be directed – this should be clarified for users.

1.10   We heard that over time, users have raised a number of requests through the group, some of whom referred to it as a 'wish list.' However, the feasibility of these requests is sometimes unclear. HMRC and ONS should be more transparent with users about the constraints of their teams and be clearer on the reasons for prioritising some developments over others. It would be helpful for HMRC to consider regularly publishing a copy of the slides used during the quarterly RTI steering group to support wider engagement in an open and transparent way.

1.11   Wider user engagement tends to be with known users or via ad hoc queries through the contact inbox. Both HMRC and ONS recognise that more could be done to harness wider user views. Broadening user reach on labour market statistics has also been highlighted as a recommendation in our review of ONS's transformation of the LFS. The development of labour market statistics in the ONS more generally, offers a good opportunity for ONS to broaden its user engagement, including users of the PAYE RTI statistics, and to develop a topic-based and cohesive labour market user engagement strategy. It would be helpful if a joined-up labour market user engagement strategy was prioritised. Support is available via the engagement hub in ONS with further resources also available on the Government Analysis Function's GSS User Support and Engagement Resource (USER) hub.

Requirement 2: To ensure that users' needs are fully understood and use of the PAYE RTI statistics is well supported:

   a) HMRC and ONS should broaden their user engagement activities to harness a wide range of user views in the ongoing development of these statistics.

   b) HMRC and ONS should communicate statistical development plans, manage user expectations about what further value can be obtained from the data and how future developments are being prioritised.

Accessibility

1.12   [Aggregate data tables](#) are published within a single Excel file each month enabling users to conduct their own analyses. We found that whilst re-use is supported, the data tables published within the one Excel spreadsheet can make navigation cumbersome and the grouping of all notes onto a summary page does not support individual use of tables across all of the separate tabs. Some users reported difficulty trying to navigate between particular tables due to the number of them – the spreadsheet published in May 2023 consists of 39 individual tabs. Providing a "quick link" back to the contents page on each tab, for example, could enable easier navigation around the spreadsheet. Understanding the user need would provide insight into any underused data tables, thereby possibly enabling rationalisation of breakdowns within the Excel sheet.

1.13   Having access to the data via an online platform such as [NOMIS](#) would enable further re-use and this was raised by users as a preferred option to complement the current Excel data tables. HMRC and ONS told us that they had considered this, and that they have no plans to add the data to NOMIS as the data are already published elsewhere. Irrespective of this, NOMIS is a widely used platform and making the data available through it would be helpful to users. HMRC told us that users can request access to de-identified microdata through its [Datalab.](#) However, the frequency and regularity of requests via this service is not known by HMRC. Some users that we spoke to were unaware of this option so HMRC and ONS should do more to publicise this service.

Requirement 3: To help enhance the value offered by the statistics by supporting users' wider analysis needs, HMRC and ONS should review the way PAYE RTI statistics are currently disseminated, and implement any improvements needed. This user focused review should include considering ways to improve navigation around the data tables within Excel, making data more widely available, for example through NOMIS and better promoting the Datalab service.

Maximising insight through data linkage

1.14   A very welcome development is ONS using data linkage to help answer questions about movement within the labour market. ONS is investigating linking the PAYE RTI data with self-assessment tax data to provide insight into the movements of people between different employment statuses, such as employment to self-employment. Users noted the potential further value of answering these questions about people moving in and out of employment and self-employment and the reasons driving the changes such as pay; and what impact government policy changes may have on these movements.

1.15   Outside of the monthly delivery of aggregate data tables from HMRC to ONS to enable the monthly publication, ONS has access via a data sharing agreement to a cut of the underlying microdata, with the latest data covering the period up until 2021, and with a further extract planned for later this year. We heard from HMRC that work is progressing across HMRC and ONS to coordinate a regular feed from HMRC to ONS but the timeline for this is still under discussion. ONS is seeking ways to exploit the microdata further to improve insight about the labour market, for example, filling data gaps on earnings: at the moment, analysing pay movement is

only possible on an annual basis using the ASHE and using the PAYE RTI data for this analysis would enable much timelier insight.

1.16   As an administrative data source with extensive coverage and granularity, there is lots of potential insight that could be drawn from this dataset to address current data gaps, for example providing breakdowns of the data by public and private sector allowing further analysis. The PAYE RTI data is already linked to the IDBR to identify the sectors that employees work in to produce monthly breakdowns by sector. The data are also linked to migrant worker data to produce breakdowns by nationality on an annual basis.

1.17   HMRC told us that developments are currently constrained by limitations of the current IT system and staff resourcing issues. We recognise these restrictions but note that in order to maximise insights from this rich data source, sufficient human, financial and technological resources need to be made available.

Requirement 4: To help maximise insight and the potential public value of the PAYE RTI data HMRC and ONS should consider ways to support user needs for additional insight by how any data gaps could be filled, whether that be through any ad-hoc analyses or additions to the regular publication if helpful and feasible.

Methodology

1.18   When the statistics are initially published, approximately 15% of the data is yet to be received. This could be due to employers filing returns late or where an employee has not been paid yet for that month's work. HMRC imputes these data. In July 2022, HMRC changed method following a review by HMRC and ONS methodologists. The changes incorporated a seasonal factor into the imputation model and made the model more responsive to recent changes to the labour market by incorporating more recent data. The aim was to reduce the scale of revisions seen in the flash estimates, which in the previous imputation method resulted in some significant revisions (over 100,000 employees) being made as more data became available over subsequent months. The magnitude and systematic bias of revisions, in particular for the flash estimate, was raised as a concern by users. Since the new method has been incorporated, revisions have become smaller and we, along with users we spoke to welcome this significant improvement. However, HMRC and ONS have not yet published any information or analysis on the impact of the new model on the magnitude of the revisions due to the need for a long enough time series to enable this analysis. We welcome HMRC and ONS's plans to do so once they have enough evidence as this will help promote confidence in the new method. It would also be helpful for users to know when to expect the analysis.

1.19   While the magnitude of the revisions has decreased, HMRC told us that it continues to assess the impact of the change to the imputation method as there may be other factors driving the decrease in revisions such as movements in the labour market being less volatile than when they were during the pandemic, which is when the original imputation method was developed. Both imputation methods are being run in parallel most months, although not as frequently as HMRC would like to due to a lack of resources. It is not clear for how long the parallel runs will continue and so plans for the parallel running of the imputation methods and any resulting analysis should be communicated to users. As the new method is bedding in, HMRC has noted that while the largest revisions to the data are still occurring in the first month,

smaller revisions are also coming through in the later months, so HMRC is also investigating the reasons for these later revisions.

1.20   In line with our expectations around communicating uncertainty, it is good to see that an explanation of the proportion of imputed data and the revisions policy are included in the monthly publication. Users appreciate the transparency of the revisions process, and the published revision triangle helps users to understand the scale of revisions made. However, some users felt that the technical details of what is causing the revisions could be clearer.

1.21   The main source of information about methods is the methodology article, jointly published by HMRC and ONS. There was a general consensus across users that we spoke to that the quality and methodology information could be improved, considering the needs of both expert and non-expert users: some users commented that information was not very accessible, and others requested the need for more technical and detailed content. In previous regulatory work we have highlighted good practice in this area such as the Department for Work and Pensions' background and methodology document for its National Insurance numbers (NINo) allocated to adult overseas nationals statistics.

Requirement 5: To support user confidence in the recent methodological changes and support ongoing understanding and development of the statistics:

    a) HMRC should publish its analysis and evaluation of the implementation of the new imputation method explaining the impact on the quality of the statistics.

    b) More widely, HMRC and ONS should update the quality and methodology documentation in ways that meet the needs of both expert and non-expert users.

Quality assurance

1.22   The HMRC statistical team told us about the quality assurance it undertakes throughout the production process once it receives a cleansed dataset provisioned by HMRC's Chief Digital and Information Officer (CDIO) team and its suppliers. This includes using thresholds at an aggregate level to identify high and low values and check the quality of the initial data cleansing. Once the aggregate tables are produced, the data are then shared with ONS, which has responsibility for publishing the statistics on its website.

1.23   At present there is no information available to users about the end-to-end quality assurance process and the level of assurance associated with the statistics. Having good links with the data providers helps to ensure continuous quality improvement and explaining assurances around quality will help reassure users and help them form an appropriate interpretation of statistical quality. OSR's review of HMRC's quality assurance processes draws out more about this, and it is good that HMRC is currently reviewing and improving quality assurance arrangements across all its statistics. Further guidance is also available on the Government Analysis Function Guidance Hub to help support producers in communicating information about quality. Our Quality assurance of administrative data (QAAD) framework is a useful tool to reassure users about the quality of the data sources.

Requirement 6: To help provide assurances around the quality of the published statistics, HMRC and ONS should publish information about the start-to-end quality assurance process for PAYE RTI statistics.

1.24   We welcome that HMRC has implemented Reproducible Analytical Pipelines (RAP) into its production of the PAYE RTI estimates. Introducing automation into the production process has helped to reduce the potential for human error and freed up some resource, allowing the team to spend more time on other aspects of the production process.

Independent decision making, leadership and orderly release

1.25   The PAYE RTI estimates are a joint publication between HMRC and ONS, with both teams' roles in the production and publication processes clearly defined internally and supported by good relationships between the two teams at a working level. HMRC supplies aggregate data to ONS for publication on its website and has overall responsibility for the PAYE RTI data. Continued effective working at all levels will be important for the ongoing success of these statistics.

1.26   ONS does not provide pre-release access to its data and HMRC publishes its list of those who do receive pre-release access. For the PAYE RTI data, HMRC reviews the list each month to ensure only those who need access to the data receive it, with any requests for additions to the list needed to be supported by a business case.