

Quality assurance of administrative data used in earnings and employment from PAYE RTI

Investigation of the administrative data used in production of ONS and HMRC earnings and employment from Pay As You Earn (PAYE) Real Time Information (RTI) data.

Contact:
ONS Labour Market team and
HMRC RTI Statistics
labour.market@ons.gov.uk,
rtistatistics.enquiries@hmrc.gov.
uk
+44 1633 455400

Release date:
10 January 2025

Next release:
To be announced

Table of contents

1. [Overview](#)
2. [Quality assurance of administrative data \(QAAD\) assessment What is a QAAD?](#)
3. [Practice area 1: operational context and administrative data collection \(QAAD Matrix score A2\)](#)
4. [Practice area 2: communication with data supply partners \(QAAD Matrix score A1\)](#)
5. [Practice area 3: quality assurance principles, standards and checks applied by data suppliers \(QAAD Matrix score A1\)](#)
6. [Practice area 4: producer's quality assurance investigations and documentation \(QAAD Matrix score A2\)](#)
7. [Summary of the strengths and limitations](#)
8. [Cite this methodology](#)

1 . Overview

Summary of the Earnings and employment from PAYE RTI publication

Each month HM Revenue and Customs (HMRC) and the Office for National Statistics (ONS) jointly release estimates of payrolled employees and their pay using HMRC's Pay As You Earn (PAYE) Real Time Information (RTI) data.

As part of RTI requirements, eligible employers must report payroll information, which includes all employees' pay, tax and deductions to HMRC each time an employee is paid. From this administrative data source estimates are calculated on the number of individuals in PAYE and the average and total earnings paid through PAYE RTI for the UK as well as split by region, age, sector and other breakdowns.

This report contains information on the RTI administrative data source, as well as a quality assessment of the source in line with the Quality Assurance of Administrative Data (QAAD) toolkit developed by the UK Statistics Authority.

2 . Quality assurance of administrative data (QAAD) assessment What is a QAAD?

The [Quality Assurance for Administrative Data \(QAAD\)](#) was introduced by the UK Statistics Authority in 2015. Under the [Code of Practice for Statistics](#), producers of statistics must review data that is used to produce statistics. The Statistics Authority encourages the use of a QAAD assessment to do this.

This report aims to apply the relevant principles of the QAAD toolkit to provide a level of assurance for the use of the [Earnings and employment from PAYE RTI publication](#). The Administrative Data Quality Assurance toolkit recognises that a proportionate level of assurance should be applied relative to the statistic's importance and some statistics need more assurance than others.

The [Administrative Data Quality Assurance toolkit \(PDF, 244KB\)](#) provides guidance for meeting assurance levels.

The assessment has been carried out in accordance with the QAAD toolkit.

The QAAD toolkit sets out four levels of quality assurance, used in the UK Statistics Authority Risk and Profile Matrix:

- A0 – no assurance: this level is not compliant with the [Code of Practice for Statistics](#)
- A1 – basic assurance: the statistical producer has reviewed and published a summary of the administrative data quality assurance (QA) arrangements
- A2 – enhanced assurance: the statistical producer has evaluated the administrative data QA arrangements and published a fuller description of the assurance
- A3 – comprehensive assurance: the statistical producer has investigated the administrative data QA arrangements, identified the results of an independent audit and published detailed documentation about the assurance and audit

UK Statistics Authority Risk and Profile Matrix

Low level of risk of quality concerns

- Lower public interest profile: statistics of lower quality concern and lower public interest [A1]
- Medium public interest profile: statistics of lower quality concern and medium public interest [A1/A2]
- Higher public interest profile: statistics of lower quality concern and higher public interest [A1/A2]

Medium level of risk of quality concerns

- Lower public interest profile: statistics of medium quality concern and lower public interest [A1/A2]
- Medium public interest profile: statistics of medium quality concern and medium public interest [A2]
- Higher public interest profile: statistics of medium quality concern and higher public interest [A2/A3]

High level of risk of quality concerns

- Lower public interest profile: statistics of higher quality concern and lower public interest [A1/A2/A3]
- Medium public interest profile: statistics of higher quality concern and medium public interest [A3]
- Higher public interest profile: statistics of higher quality concern and higher public interest [A3]

The QAAD toolkit outlines four specific areas for assurance, and the rest of this report will focus on these areas in turn. These are:

- operational context and administrative data collection (Section 3)
- communication with data supply partners (Section 4)
- quality assurance principles, standards and checks applied by data suppliers (Section 5)
- producer's quality assurance investigations and documentation (Section 6)

Each of the four practice areas are evaluated separately, and the respective level of assurance is stated.

3 . Practice area 1: operational context and administrative data collection (QAAD Matrix score A2)

Definition

This relates to the need for statistical producers to gain an understanding of the environment and processes in which the administrative data are being compiled and the factors that might increase the risks to the quality of the administrative data.

Background information on PAYE

HM Revenue and Customs (HMRC) is the department of the UK government that is responsible for the collection of taxes, including Pay As You Earn (PAYE). In 1944 the PAYE system was introduced whereby tax was deducted from wages by employers during each pay cycle, for example, each week or month.

PAYE is a tax on all payments of wages and salary or other compensation such as sick pay, maternity pay, directors' fees, and pensions. It is deducted by the employer from those payments where certain criteria are met. Specifically, if any employee is paid above the National Insurance Lower Earnings Level (LEL), get expenses and /or benefits, have another job or get a pension, then the employer is responsible for reporting the payments and deductions made and for sending the tax on to HMRC each month.

Information on RTI

Real Time Information (RTI) was an important government programme to improve the way in which employers submit PAYE National Insurance contributions (NICs) and Statutory Payments information to HMRC. Previously employers and pension providers sent information about tax, NICs and other payroll deductions to HMRC after the end of each tax year. However, under RTI, employers and pension providers send HMRC data online in real time about tax, NICs and other deductions when or before each salary or earnings, or pension payment is made.

RTI makes the PAYE process simpler and less burdensome for employers and HMRC, for example, by removing the need for the end of year return (P35 and P14) and simplifying the employee starting and leaving processes. It also makes PAYE more accurate for individuals, over time reducing the number of bills and repayments sent after the end of the tax year.

RTI was first introduced in April 2012, with a pilot service for volunteer software developers and employers over a period of 12 months. Following on from the pilot, most employers started reporting their PAYE information in real time from April 2013, with a number of non-standard PAYE schemes finally starting to send their PAYE information in real time in April 2014. Since 6 April 2013, all new PAYE registrations are required to join RTI unless they have an accepted claim for exemption from online filing.

Employer responsibilities and information submitted

Each time an eligible employer pays their employees they must submit a Full Payment Submission (FPS) to HMRC. The FPS is the main RTI electronic submission which an employer completes and submits to HMRC each time that they pay their employees, regardless of the expected length of employment or amount of pay, to advise HMRC which employees they have paid and to give full details of the payment and the deductions from it.

This gives details of the tax, National Insurance and student or postgraduate loan deductions they have deducted, which they need to pay over to HMRC. It also provides details of any Statutory Maternity Pay (SMP), Statutory Adoption Pay (SAP), Ordinary Statutory Paternity Pay (OSPP), Statutory Shared Parental Pay (ShPP) or Statutory Parental Bereavement Pay (SPBP) they have paid out to employees. Details of any Statutory Sick Pay (SSP) paid to employees are only required if the employer is able to make a claim for recovery in respect of that SSP. These data will be used to calculate how much an employer must pay to HMRC after the end of each tax month or tax quarter (if paying quarterly).

Employers must report all payroll information to HMRC each time they pay an employee. This must include payments on standard paydays, and any additional payments or amounts recovered from the employee. Eligible employers need to report the payment details of all employees paid no matter how much they are paid, even those earning below the Lower Earnings Limit (LEL), or those paid just once a year. If an employer has not paid any employees in a pay period and has no returns to make, they need to inform HMRC of this using an Employer Payment Summary (EPS).

RTI legislation states that an employer is expected to report payroll information on or before the date the individual is paid. If an employer fails to do this, they may be charged a filing penalty.

Other information sources used

These RTI submissions are the main source of data and are the only data used to form estimates of the number of people employed and their earnings. However, in the publication we split estimates out into demographic groups, including by industry sector. Business activity is collected through RTI, but it is not compulsory to provide and often cannot be meaningfully categorised into consistent groups. Therefore, we supplement the RTI data with two other sources to assign Standard Industrial Classifications (SIC 2007) to employers.

The Inter-Departmental Business Register (IDBR)

This is a comprehensive list of UK businesses used by government for statistical purposes. The IDBR provides the main sampling frame for surveys of businesses carried out by the Office for National Statistics (ONS) and other government departments. It is also an important data source for analyses of business activities. This database system holds information, which can be used to identify the main characteristics about employers in our data. The publication team uses the IDBR to collect additional information on employers not contained within RTI.

Companies House

In cases where industry codes cannot be provided through the IDBR, then additional codes can be added via text-matching to Companies House data. Companies House is the executive agency of the UK government that maintains the register of companies, employs the company registrars and is responsible for incorporating all forms of companies in the UK. The data used in this release are drawn from a publicly available monthly download, which is then matched to PAYE RTI data via company name.

For a minority of cases, no SIC 2007 code can be provided even with these extra steps. For these cases where possible we impute a SIC 2007 code based on employer characteristics and other demographic information.

Strengths

- RTI data cover the full payroll population rather than a sample of people or companies; this allows detailed breakdowns of the population.
- Eligible employers are required by law to provide accurate and timely returns.
- The data are updated in real-time, with the analytical data tables used to produce the statistics created daily.
- Returns are made through software tested to be safe and secure.
- Guidance is given to employers and software providers to ensure returns are accurate.
- Employers can correct errors themselves.

Limitations

- The PAYE RTI data exclude some individuals which other sources could class as employed, such as some self-employed individuals or employed individuals in the undeclared economy.
- Data are on a payments basis and need to be calendarised to the period worked, which affects the timeliness of the data.
- Only the data required to operate the PAYE tax system are reported on RTI, so not all details on employments are available (for example, occupation).
- As with all admin data, PAYE RTI carries a risk of inputting errors or fraudulent returns.

4 . Practice area 2: communication with data supply partners (QAAD Matrix score A1)

Definition

This relates to the need for data suppliers to have collaborative relationships with the stakeholders involved, which outline any formal agreements, and/or engagements with suppliers and users.

The [Earnings and employment from Pay As You Earn \(PAYE\) Real Time Information \(RTI\) publication](#) is jointly published by HM Revenue and Customs (HMRC) and the Office for National Statistics (ONS), using the RTI data collected by HMRC as outlined previously.

From the collected raw data a dataset is created for HMRC analysts to query and extract data from. This is used by the HMRC RTI publication team to create the aggregated outputs used within the bulletin and published in the supporting data tables, which are then passed to the ONS for publication. With a number of teams involved in the collection of the RTI data, the preparation of the aggregate tables and their publication and usage, there are a number of means of communication.

Communication with employers

The RTI data provided to HMRC are directly used in the calculation of Income Tax and to ensure payments made by employers are both correct and kept up to date. HMRC provides a number of different ways to support employers to ensure that the data provided to HMRC are as accurate as possible. Guidance is published online advising employers what data items they need to provide, where to include that information and how often they need to make a submission. There is also online support to software providers, describing specifications and guidelines to aid designing software for RTI submissions. Employer bulletins are also sent out regularly advising employers of issues and upcoming changes.

The Employment and Payroll Group is HMRC's principal forum for HMRC and other government departments to engage with the employment and payroll community. HMRC uses this to explain and explore implications of potential changes to policies, products and processes affecting employers and intermediaries, and provide an opportunity for early review of guidance and information to make sure processes are clearly explained so that customers understand what is required of them.

HMRC also works closely with large businesses and consults with them in various forums, including the [Business Tax Forum](#) and the [Large Business Customer Survey](#).

Internal HMRC communication

Within HMRC there are a range of teams who use RTI data for operational and analytical purposes. A number of networks have been set up to share knowledge and raise issues regarding RTI data in the hope of raising quality standards and raising data issues quicker.

Within HMRC's analytical community there are also initiatives driving cross-team quality assurance and other forms of peer-review. These have been used to assess the data and methods used by the RTI statistics production team.

Communication between the ONS and HMRC

In the week leading up to publication each month there is a meeting between the HMRC and ONS publication teams as well as other ONS labour market statistics producers. Within this meeting we discuss changing trends, quality concerns, comparisons with other publications and any other issues affecting that month's published data. This publication and underlying data feed into a range of other sources and publications, therefore we have ongoing communication with a range of stakeholders.

Throughout the year the publication teams across HMRC and ONS adopt a collaborative approach to working, with regular contacts to discuss workplans and potential methodological changes.

Communication with publication users

We continue to seek feedback from users of the publication both internal to government and external. Once a year or more we conduct a steering group with government users of the statistics, informing them of upcoming changes and collecting suggestions for further improvements to the publication. In recent years this has led to changes to the publication to ensure the users get what is most useful for them, including expanding the number of tables and breakdowns we provide to include geographies, age, industry sectors and combinations of these breakdowns. We closely work with stakeholders such as HM Treasury, the Office for Budget Responsibility (OBR) and the Bank of England, as well as regional statistical teams.

We also publish contact details within each publication bulletin, and encourage users to contact us with requests, suggestions and any quality issues they may have regarding the publication.

Strengths

- Employers are provided guidance and support to fill in submissions correctly and are kept up to date of upcoming changes to the system.
- PAYE RTI data users within HMRC communicate regularly via email, meetings and user group and forum announcements and discussions.
- A collaboration between HMRC and the ONS has been drawn up in line with the principles and protocols of the [Code of Practice for Statistics](#).

Limitations

- Further improvements could be made to communication between HMRC operational and analytical teams to encourage sharing of information and raising data concerns more quickly; work is currently under way to improve these networks and set up closer collaboration across these teams.

5 . Practice area 3: quality assurance principles, standards and checks applied by data suppliers (QAAD Matrix score A1)

Definition

This relates to the procedures in place for data collection, quality and inspection process from suppliers through the process before it reaches the statistics publication team.

HMRC quality assurance checks on the PAYE RTI

HM Revenue and Customs (HMRC) is the UK's tax, payments and customs authority and has a vital purpose: they collect the money that pays for the UK's public services and give financial support to people. It is HMRC's job to make it easy for customers to get tax right, and hard for anyone to bend or break the rules.

As part of this, HMRC conducts compliance checks on Pay As You Earn (PAYE) records and returns to ensure that employers are reporting earnings correctly and therefore paying the right amount of tax. Where these checks identify under- or over-paid tax, records are subsequently amended and, where appropriate, penalties applied.

The tax gap is the difference between the amount of tax that should be paid in theory and what is actually paid. HMRC's [Measuring tax gaps publication](#) provides details on the current tax gaps and on steps being taken to reduce them.

Tax gap figures for the year 2024 show that the PAYE employer compliance tax gap is 0.9% of the theoretical PAYE tax liability in the 2022 to 2023 tax year. This has followed an overall reduction from 1.6% in the 2005 to 2006 tax year. This supports the view that the activity that HMRC does for non-compliance, and ensuring taxpayers are paying the correct amount of tax is effective, but also suggests that the majority of the data collected through PAYE RTI are accurate.

However, it is recognised that errors can be made in reporting, as well as some evasion, fraud or activity in the hidden economy. While tax gap estimates show that this is likely being picked up by HMRC's compliance activity, it can take time for records to be subsequently corrected, and this is reflected by the level of revisions in the RTI employment publication. Therefore, although we deem the RTI admin data source to be of high quality, there is a difference across time, with more recent data being of slightly lower quality.

Strengths

- Many teams conduct their own quality assurance checks, which allows various teams to pick up any errors in the data source.
- Compliance within PAYE is generally high, signalling that the quality of the administrative data source is high.

Limitations

- The most recent data are more likely to contain errors or incorrect reporting, making it of slightly lower quality.

6 . Practice area 4: producer's quality assurance investigations and documentation (QAAD Matrix score A2)

Definition

Producer's quality assurance (QA) relates to the quality assurance conducted by the statistical producer, including collaboration against other data sources.

As outlined in Section 4: Communication with data supply partners, after being collected the Pay As You Earn (PAYE) Real Time Information (RTI) data are transferred to an analytical data repository overnight and are stored in a format from which they can be queried for management information and policy analysis. This forms the source of the RTI data processed and used to generate the statistics.

Where there are errors found, employers can submit corrected returns with revised values. If there is fraud or a failure to take reasonable care, HM Revenue and Customs (HMRC) compliance teams can investigate and this may lead to further corrections submitted. These revisions will be transferred into the analytical data in the same way.

However, recent months are reported in the statistics to give as timely a view as possible, and it is likely that the data are processed for the statistics with some of these early errors and inaccuracies still included. To account for this, we do the following.

Firstly, revise the historical series by re-extracting the data each month for the current tax year and the previous tax year. This ensures that changes and amendments are incorporated into the data as soon as is practical. We also then once a year re-extract the data for the entirety of the time series to collect any revisions that affect earlier years. This balances out the need to collect all revised data points against the increased processing time from running the whole series.

Secondly, treat the entire data series with the potential to have errors and conduct quality checks on the entirety of the data each month. Therefore, although we deem the admin data source to be of high quality, we conduct quality assurance to a level necessary as if it were of lower quality.

We also maintain a number of steps and checks to ensure that the data quality is maintained through its processing and movement. These steps and quality checks are now described.

Reproducible Analytical Pipeline

The RTI data are accessed, extracted and processed using SAS software. Aggregated data is then also further processed using R and markdown code. The process and code have been developed into a Reproducible Analytical Pipeline (RAP).

As part of this, the code has been written to automate as many steps as possible, minimising any manual edits to data or outputs. The code is written so that where changes are necessary to produce the statistics, for example, to change date parameters, these are changed in one place and then passed through the remainder of the code automatically. For the main parts of the code, these updates are checked by a second person to make sure the code is running the correct parameters.

If changes need to be made to the code structure or the automated processes, these are checked by a second person. If deemed large enough, the changes are run in parallel to compare with current outputs before being run for the publication itself. Version control is maintained so that older versions can be accessed and put in place if an issue is found.

Transformations to the data

The PAYE RTI data are collected and stored on a payment's basis, where records are stored for each submission made to HMRC. Summing payment amounts and the number of records to produce monthly estimates for the total pay earned and number of employees in work would be problematic, especially when aggregating across both jobs paid monthly and jobs paid weekly as the number of weeks in a month can vary.

A calendarisation method converts these payments data towards a dataset of jobs and pay rates from which monthly employment counts and pay estimates can be calculated. These estimates are then more comparable with other labour market sources and UK National Accounts. This also reduces the volatility of the time series produced using RTI data.

Adjustments to the data are also made for incorrectly reported payment frequencies and missed and double payments, by looking at the historical record for each employment and adjusting any inconsistencies.

To increase the accuracy of recent estimates, imputation methods are applied for missing records, using probabilistic imputation methods to determine employment terminations or new hires where data is missing.

More information on these transformation methods can be found in our [Monthly earnings and employment estimates from Pay As You Earn Real Time Information \(PAYE RTI\) data: methods](#) publication.

Data processing checks

As a first step we check payment submissions to make sure they can be uniquely identified and linked to an individual. In some instances it can be possible to pay tax in the UK using Temporary Reference Numbers (TRNs) while waiting for a National Insurance number (NINo). We check whether the proportion of payments belonging to TRNs as opposed to NINos has changed compared with historical proportions.

Payments from pensions schemes are also made through the PAYE system and are contained within the data. As we only include employees and their earnings within the statistics, pension payments are identified and removed. The number of new pension payments and schemes added each month is compared against expected boundaries. If monthly changes are higher or lower than expectations, schemes or individuals are investigated to assess if there is a problem with the flag criteria.

Logs from running code are checked for errors or code bugs. The logs are checked through the processing stages to make sure that the population is being maintained throughout the code, and any changes in the size of the dataset reflect intentional additions or removals. Similarly, run-times are checked against expected boundaries, which could indicate issues affecting the data quality.

Individuals with the highest levels of pay and highest monthly changes in pay are flagged, and where these measures are outside expected norms these individuals or schemes are investigated to make sure the data are correctly reflecting a change in circumstance or other valid reason for pay changes.

Output validation

Checks are carried out to ensure that the produced outputs match the datasets that feed into them. The aggregated outputs tables are checked to make sure there are no missing values. Totals are compared back against previous levels, comparisons being made for figures at the UK level, the distribution of pay and median pay growth rates, as well as trends across different pay measures.

Following seasonal adjustment, both the seasonally adjusted and the non-seasonally adjusted data for each published series are visually compared with the previous month's estimates, to ensure revisions are realistic and are not being artificially changed in one of the processing stages.

Final outputs are sent from HMRC to the Office for National Statistics (ONS) Labour Market team, who check outputs against other data sources to check coherence. Outputs are then presented to senior colleagues at HMRC and the ONS so that trends or inconsistencies can be discussed and investigated. Unusual patterns are identified and investigated to see if these are because of the methodology or reflect changes in the labour market.

While some of these checks are still currently manual, we are still working to improve the RAP by introducing unit testing and further automating as many checks as possible.

Strengths

- The statistical production team conduct a wide range of QA checks on the received admin data.
- Code is designed according to RAP principle, with further plans to improve this.
- Outputs are closely examined to check for issues caused by processing.
- A range of analytical experts review the outputs to check trends are realistic and identify areas for investigation.
- Calendarisation methods convert the payments data to a jobs dataset comparable with other labour market publication and international guidelines.
- Imputation methods allow timelier estimates of recent months.
- Methodological changes are reviewed across multiple teams, with independent peer review being made for the most important changes.

Limitations

- The process of historical records being corrected and updated increases the number and scale of revisions to the data each month.
- Imputation methods also increase the frequency and scale of revisions to recent data.
- The size and nature of the RTI data as well as the complexity of the extraction and transformation processes increase the risk of data being manipulated through processing, increasing the need for careful QA and applying processing according to RAP principles.

7 . Summary of the strengths and limitations

HM Revenue and Customs (HMRC) and the Office for National Statistics (ONS) consider the main strengths and limitations of the Pay As You Earn (PAYE) Real Time Information (RTI) data for its purpose to be as follows.

Strengths

- RTI data used in the publication cover the full payroll population rather than a sample of people or companies; it refreshes data on a monthly basis, which allows us to provide detailed breakdowns of the population.
- It is a monthly publication and offers an earlier estimate on labour market conditions than other publications.
- Eligible employers are required by law to provide accurate and timely returns, with data then being updated in real-time.
- The team carries out extensive quality assurance throughout the statistical production process.

Limitations

- As with all admin data, PAYE RTI data carry a risk of inputting errors or fraudulent returns.
- The data capture and processing involves a range of teams, which include operational and compliance teams; therefore the quality of the statistics is dependent on collaboration and communication across a number of teams.
- The complexity of transformations necessary to produce published estimates increase the need for careful quality assurance of the data and its processing.

Earnings and employment from PAYE RTI statistics has been assessed to being "enhanced assurance" as outlined by the UK Statistics Authority QAAD toolkit.

We will be taking next steps to investigate the limitations outlined in the various practice sections and these will be communicated to users in future quality assurance of administrative data (QAAD) report updates.

If you are of the view that this report does not adequately provide this level of assurance, or you have any other feedback, please contact us via email at rtistatistics.enquiries@hmrc.gov.uk with your concerns.

8 . Cite this methodology

Office for National Statistics (ONS), released 10 January 2025, ONS website, methodology, [Quality assurance of administrative data used in earnings and employment from PAYE RTI](#)