

# Owner occupiers' housing costs in CPIH: quality assurance of administrative data, Dec 2016

CPIH is a version of CPI that includes the cost of owning, maintaining and living in one's home, this element is known as the owner occupied housing cost (OOH). This paper aims to inform users of the quality of administrative data used in OOH, due to its high impact on CPIH.

Contact:  
Jonathan Lewis  
jonathan.lewis@ons.gsi.gov.uk

Release date:  
6 December 2016

Next release:  
To be announced

## Table of contents

1. [Executive summary](#)
2. [Introduction](#)
3. [Private rents data overview](#)
4. [Valuation Office Agency](#)
5. [Welsh government](#)
6. [Scottish government](#)
7. [Annex A: New quality assurance practices for private rental data](#)
8. [Annex B: Flow diagrams of private rents quality assurance processes](#)

# 1 . Executive summary

The following are the main points from this report.

CPIH is a version of CPI, currently the main measure of inflation, which includes the cost of owning, maintaining and living in one's own home. This element is known as the owner occupied housing cost (OOH).

CPIH was a National Statistic until 2014, when it lost its status due to concerns over the underlying data.

There are 4 sources of data that are used in the calculation of OOH; administrative data is used for England, Scotland and Wales, whilst Northern Ireland currently uses price collection.

Administrative data is not sourced using statistical methods and therefore requires rigorous quality assurance to ensure that they are suitable for use in a National Statistic.

We do not have full access to the microdata for the Valuation Office Agency, therefore this is seen as higher risk and has been subject to far more rigorous scrutiny and assurance than other administrative data sources. This includes an annual audit and the publication of a report comparing it with other sources.

Numerous new quality assurance measures have been undertaken on private rents data since CPIH was de-designated as a National Statistic. A summary of these can be found in Annex A.

These extra measures have instilled us with confidence in the quality of the underlying data and its aggregation. We have established processes to monitor OOH to ensure the quality of and confidence in the statistic remains high.

This report has been released to assure users of the quality of administrative data used in OOH, due to the high impact it has on CPIH. A further document, which assesses all data sources used in the construction of CPIH against the UK Statistical Authority's quality assurance of administrative data (QAAD) toolkit, will be released at a later date.

## 2 . Introduction

The Consumer Prices Index (CPI) is a high-profile statistic that is used as our headline measure of inflation. It is used by the Monetary Policy Committee for inflation targeting and to set monetary policy. CPI including owner occupiers' housing costs (CPIH) is identical in construction to CPI, but includes the costs associated with owning, maintaining and living in one's own home. It is our intention that CPIH should become the headline measure of inflation from March 2017, as the cost of owning a home is an important component of household expenditure. There is therefore a large level of public interest in, and scrutiny of, CPI and CPIH.

CPIH was launched in 2013 and in November 2013, CPIH was designated as a National Statistic. In August 2014, however, the [National Statistics](#) status of CPIH was discontinued after issues emerged with its "live running" of CPIH, relating to how it processed Valuation Office Agency (VOA) data to estimate owner occupiers' housing costs (OOH). In response to the high level of public interest in and scrutiny of these issues, we have since taken steps to put in place new quality assurance processes and to instill confidence in the procedures we have in place for all underlying data in CPIH and in particular the administrative data we use from the VOA. To demonstrate that the statistic is reliable and meets the requirements of National Statistic status, a summary of improvements made to the quality assurance process can be found in Annex A.

After improving our quality assurance procedures, we have high confidence in the data sources used in the construction of CPIH as a whole. Some administrative data sources have areas for improvement, for instance in improving the communication between ourselves and data suppliers, and reviewing Service Level Agreements with suppliers to ensure that we are made aware of any changes in data collection or quality assurance producers by the supplier. For these reasons and due to the importance of the statistic, all administrative data sources have been subject to an enhanced quality assurance assessment. We will in the future be releasing a full assessment of all data sources used in the construction of CPIH, using the UK Statistical Authority's quality assurance of administrative data ([QAAD](#)) [toolkit](#).

CPIH uses administrative data from the VOA, comparative data from the Welsh and Scottish governments and additional data from the local CPI collection for Northern Ireland, to calculate its rental indices as a proxy for the cost of owning one's own home. For more information on the methodology used, please see our [CPIH Compendium](#).

OOH are not easy to measure and given the high weight attached, the methodology used to construct OOH has been the source of much debate and there is a higher degree of scrutiny of its data and methodology than other elements of the CPIH basket. This article therefore provides users with a comprehensive analysis of the current quality assurance procedures undertaken on VOA, Welsh government and Scottish government administrative data by us and the data suppliers separately from the forthcoming full CPIH assessment; we have also assessed our communication with the suppliers and detailed both improvements that have been made and areas for future improvement. The aim of this report is to reassure users of the quality of the OOH data.

## 3 . Private rents data overview

### 3.1 What is administrative data?

Administrative data is that which, unlike surveys such as the Living Costs and Food Survey (LCF), is not originally created for a statistical purpose and is often sourced from other government departments or a third party. As such there is a higher level of risk in its use and it is subject to a higher level of scrutiny and assurance before being used in the construction of official statistics.

Administrative data from the Valuation Office Agency (VOA), Welsh and Scottish governments are also used in the construction of the Index of Private Rental House Prices (IPHRP). This is an experimental price index for rental prices, which uses the same data as OOH, but uses different methodology and weights to reflect the private sector rather than the owner occupied sector.

There is an information management and governance structure in place to manage the use of administrative data sources. CPIH is constructed by two teams which sit within Prices Division: Production and User Engagement team, and Production and Commodity Analysis team. The Prices operations and management board discusses quality and other methodological issues. Prices Division sits under the National Accounts and Economic Statistics Directorate, and seeks advice on all aspects of consumer price indices from the Advisory panels on Consumer Prices. The Technical panel is chaired by the Director of NAES, and the Stakeholder panel by an independent expert. These provide independent advice to the National Statistician on the users and applications of consumer price indices to ensure these statistics meet the needs of users. The Board of the UK Statistics Authority then makes decisions on major changes to consumer price statistics on the advice of the National Statistician. A visual overview of the structure can be found in figure 5 of [Annex B: Flow Diagrams of private rents quality assurance processes](#).

### 3.2 Measuring private rents in CPIH

The OOH component of CPIH is measured using a method called rental equivalence. This uses the rental market as a proxy for home ownership costs and asks the question "How much would I have to pay in rent to live in a home like mine?". These rental costs are obtained from VOA, Welsh and Scottish governments.

Following the de-designation of CPIH as a National Statistic due to issues emerging with its “live running”, we reviewed the series alongside VOA and identified shortcomings with the processing and implementation of the methodology. These [improvements and revisions](#) were implemented in March 2015. Similar improvements to these were made to the Welsh government and Scottish government data.

Legislation governing the VOA means that we do not have access to VOA microdata, so we have worked with VOA staff to implement improved methodology and validation procedures in their systems. We have also continued to work closely with analysts in the VOA, publishing [additional analysis](#) to explain the differences between our rental measures.

Given that much of the rental data processing is implemented by VOA and subsequently provided to us, the administrative data is deemed to be a higher risk to the quality of CPIH and a more comprehensive level of quality assurance to other data sources used in the construction of the index is required. Rental data for England, provided by the VOA, accounts for around 20% of the weight in CPIH and as recommended by the [Johnson Review](#), the index will become the headline measure of inflation in March 2017. For more information on weights and how they affect CPI, please see the report [Consumer Price Inflation: 2016 Weights](#).

We have access to both Welsh and Scottish government microdata and as there is a good appreciation of the context in which the data are collected with the application of the methodology being implemented internally by us, this represents a lower risk of quality concern than VOA. Rental data from the Welsh government accounts for around 4% of the OOH element (less than 1% of CPIH), while rental data from the Scottish government accounts for around 7% of the OOH element (around 1.5% of CPIH). Northern Ireland rental indices are not currently created using administrative data and so will not be considered in this report. We are currently investigating an opportunity to use administrative data from the Northern Ireland Housing Executive (NIHE). This is reflected in our work programme.

## 4 . Valuation Office Agency

### 4.1 VOA data collection and context

Valuation Office Agency (VOA) rent officers compile and maintain a database of private market rents in England, which is a reliable source of administrative data. Prices are collected by rent officers from landlords, letting agents and tenants, with the aim to collect approximately 15% of data from sources other than letting agents. A Central Support Unit uses tools to take an overview and independently identify potential collection issues for attention. They track collection against the overall aim, agreed with the Department for Work and Pensions, for a sufficient sample to comprise 10% of the private rental market. The English Housing Survey provides further evidence of continuing growth of the private rental market.

The criteria within the legislation that shape the rental data collection have remained consistent since the implementation of Local Housing Allowances (LHA) in 2008 and continue into the government’s Universal Credit arrangements. Although there is no sampling frame with which to establish a statistical sample, planning and monitoring tools using the 2011 Census data are used by the rent officers. They provide a means to track collection of property types and distribution against the census baseline, highlighting potential issues for attention and helping with resource allocation. The rent officers are required to interpret the available information on the market and use their judgement on the sufficiency of the data collected “to enable a local housing allowance to be determined which is representative of the rents that a landlord might reasonably be expected to obtain in that area”. With an absence of definitive metrics of the size and distribution of rental data, the rent officer’s judgement helps to mitigate the risk of an unrepresentative sample.

The risk of data provision dropping significantly due to the withdrawal of co-operation by data providers, which is caused by rents being provided on a voluntary basis, is partially mitigated by the wide range of data providers. Loss of larger lettings agents, which provide the greatest volumes of data, presents a risk to the continuity of the composition, distribution and volumes of data. Experience shows that the issues most likely to affect their decision-making are their degree of confidence in the protection of their data and internal security compliance, IT issues (withdrawal of relevant software reports over system efficiency issues) and the amount of time it takes to manually collect at the local branch level. The VOA engage with the provider at the highest level as soon as issues emerge. Most have shown they will compromise to continue with data provision until their issues are resolved.

There are ongoing restrictions on the level of data that can be shared with us; however, the VOA has worked with us to answer questions, understand and explain changes identified in the data. For example, changes in Housing Benefit following the cap introduced to Local Housing Allowance in April 2011 may have influenced a shift in rental data, with movement in the sample from Outer to Inner London [Explaining private rental growth](#).

Within consumer price statistics the usual practice is to attempt to follow the same item over time. In a rents context, the aim would be to follow the same properties over time. Given the VOA dataset is an administrative data source there is no guarantee that an updated price will be received for a property. Analysing the properties that made up the 2013 sample, nearly 40% received a price update. The remaining records were replaced within the sample through comparable or non-comparable replacements consistent with consumer prices methodology for other items.

The most significant factor determining how the data is collected is that there is a lack of any legal obligation for institutions such as estate agents or landlords to share rental data with the rent officer. All the data shared is as a result of the goodwill and trust the rent officers have established with data providers. There are no data agreements in place and no data is paid for. The logistics of implementing and actively managing such arrangements with so many potential individual sources of data are considered both prohibitive and unnecessary as voluntary data provision has consistently resulted in an annual data sample of around 500,000 rents since 2009.

The main statutory duties that shape the rent officer data requirements are the following.

Housing Benefit, Local Housing Allowance and Universal Credit

Rent officers (Housing Benefit functions) Order 1997 as amended

Rent officers (Universal Credit functions) Order 2013 as amended

Fair rents

Rent Act 1977

For the rent officer's purposes, rental data is defined as the rent paid for the tenancy. An advertised rent does not meet the criteria because a tenancy does not exist. Advertised rents are proposed rents, subject to negotiation. The degree of negotiation if any, up or down, depends on market conditions and therefore fluctuates geographically, by property type and over time.

The rental data comprises rents agreed as a result of a letting to a new tenant, a renewal agreement with an existing tenant, or a rent increase during a statutory periodic tenancy. The existing database does not allow these data to be categorised or analysed. However, the collection processes ensure that every effort is made to capture the renewals and rent increases so that the data represents both the flow and stock of the market (the "rents payable"). The stock of the market is those properties already being rented as part of a tenancy agreement. The flow of the market is the measure of new properties coming onto the market.

Rent officers are directed by law to “assume that no-one who would have been entitled to Housing Benefit had sought or is seeking the tenancy” when making their Housing Benefit-related determinations. So a tenancy identified as supported either wholly or partly by housing benefit is not knowingly added to the Fair Rents database. Local authorities are legally obliged to provide the rent officer with information pertaining to Local Housing Allowance claims to assist with data validation.

Monitoring tools provide a range of views of ongoing collection against the census figures and distribution. These include checks against volumes by category and property at different geographical output levels.

Statistics or a full dataset of Housing Benefit claims at medium super output area geographical level by property type and size are not available, so rent officers have to apply their local knowledge. The distribution is not even and it can comprise a very high percentage of the available private rental market stock in any one area, making a 10% sample difficult to achieve.

The composition of the rental data collected relates directly to the statutory purpose it supports. The statutory duties of the rent officer do not require the data to be a statistical sample. It is a purposive sample forming an administrative dataset. The collection methodology, which is more of a framework, has evolved to satisfy the rent officer's data requirements to fulfil their statutory duties.

There is strong confidence in the core data attributes used for Local Housing Allowance, Universal Credit and our statistical work. These are: address, tenancy start date – month and year (90% plus capture), property type, furniture, number of bedrooms, type of tenancy, rent payable and period and services.

Other data attributes are collected to support comparable valuations for Housing Benefit and Fair Rent purposes and to enrich the dataset and understanding of the market: age, number of living rooms, kitchen, bathrooms, condition and free text comments. It is not feasible or necessary for every record to include these, and some – such as condition – are subjective and relative to the location.

There are 6 main collection methods used by rent officers.

Visits or phone calls to lettings agents and corporate landlords

A short structured interview with the source confirming rents for properties that have recently been let. Generally the rent officer will use recent lettings lists as a prompt for the agent and, where appropriate, details of records currently on our database, which need to be updated following a new re-let or a renewal or rent increase with the same tenant. This process allows the rent officer to query any unusually high or low rents and to confirm additional details, such as the tenancy start date, as part of the dialogue. The rent officer will seek insight into what is affecting changes.

Reports from letting agent or management company software system

Many larger businesses use one of a number of standard property management software systems which can be used to generate a report detailing all lettings made within a defined period. Once the relevant template is set up this method allows rent officers to collect all the lettings an agent has completed each quarter, which effectively guarantees the volume of data collected from the source. High-volume digital collection has been piloted but requires further development to enhance the data processing aspects to accommodate the great variance of how the data is held by providers.

Tenant surveys

In locations where definite tenant groups exist (such as universities or large employers), rent officers conduct surveys either in the workplace or on the street where short standardised interviews take place with tenants about accommodation.

## Telephone enquiries or survey by correspondence

Using online or newspaper private advertisements, rent officers may make contact with the landlord direct to confirm the agreed rent and details of the property. As a general rule this type of activity tends to focus on room lettings, which reflects the more casual and fragmented nature of this market.

## Landlord correspondence

Contacting a list of known landlords each year to establish or update the achieved rents for their properties and reply via a Freepost envelope. Using this process, rent officers contact in excess of 50,000 landlords annually with a response rate of between 5 and 10% nationally.

## Trade events

Rent officers attend a range of landlord, investor and letting agent trade events; these range from local authority landlord forums to national exhibitions and professional conferences with up to 1,000 delegates. The emphasis here is to proactively engage with as many important influencers and potential influencers or advocates as possible, with the aim of generating new contacts and maintaining the VOA's profile within the private rental market community.

Rent officers validate all data collected using the methods we've mentioned by following up with sources where there is any question relating to rent, property attributes or validity. Ultimately, data is only used when the rent officer is satisfied it is genuine.

Rent officers are expected to maintain a high standard of knowledge of the private rental market in their area and over time the collection is refined using local market knowledge to reflect the changing rental market. Where necessary, resource is diverted from the regular programme of data collection to address any perceived area of weakness in the data. Rent officers follow regular collection cycles focusing on face-to-face contact with sources; where practical, they aim to replenish data before it reaches the end of its 12-month lifespan.

There are factors that can make replenishing existing data difficult:

- a landlord can change their letting agent between tenants
- they can decide to let privately rather than through the agent the data were collected from by the rent officer
- they may let through an agent on a "find only" basis then self-manage the tenancy throughout the occupation of the tenant with no further involvement of the agent
- the property may be removed from the private rental market or changed from a family let to a house in multiple occupation (shared) letting
- the letting agent may deal with new lettings but manage the renewals centrally on a different IT system or even sub-contract the ongoing management (may result in renewals not being captured)
- up to 10% of the data provided does not contain sufficient address details to make an exact match when the data comes in again (there is still a degree of reluctance to share the full address, particularly in London)

Rent officers are aware of the potential replenishment issues so explore solutions with data providers.

## 4.2 How we communicate with VOA

The VOA has and continues to work closely with us to ensure the efficient and accurate data collection of the target sample size and explore solutions to the restrictions that legislation places on the sharing of rental data. Due to the data being provided on a trust and goodwill basis, any proposed data sharing with us would need to be the subject of a consultation and impact analysis exercise to ensure it would not affect the providers' willingness to co-operate and subsequent incoming data.

VOA statisticians and rent officer management have set up regular liaison internally to identify and resolve emerging collection or data issues, checking understanding of the data and market behaviours and improving the planning aspects of data collection. A formal Operating Level Agreement (OLA) is in place between the VOA's statisticians from their Information and Analysis directorate, rent officers from the Housing Allowances team and IT support services from Digital Support. This OLA is reviewed every 6 months.

Service Level Agreements (SLAs) between ourselves and the VOA document activities such as data requirements, data transfer process and data protection. A delivery schedule is included within an appendix of this SLA which is reviewed annually. This schedule has always been met. Under this SLA, we require 3 months' notice of any changes in processing code, data collection or format of data delivery to ensure sufficient time for amendments, testing and sign-off.

Agreements are reviewed annually to see if the aims, benefits and terms of the SLA are mutually acceptable or require some adjustments. Regular meetings between ourselves and the VOA are timetabled to discuss performance and future plans. The meetings follow this structure:

- monthly meetings over teleconference between operational contacts
- quarterly meetings (teleconference or face-to-face) between operational and Grade 6 contacts
- half-yearly face-to-face meetings between operational, Grade 6 contacts and Information Asset Owners

We have 2 members on the VOA's Peer Review Group meeting, which focuses on the quality assurance process for the development of the VOA's own private rental market statistics. As a user of VOA statistics, we are also a member of their Domestic Statistics Advisory Panel, which meets twice a year.

As well as these formal arrangements there are also meetings held as necessary when specific issues arise which need addressing.

## 4.3 Quality assurance principles, standards and checks by data suppliers

There are a range of quality assurance and data validation processes that take place, either during the collection process as the data is entered onto the VOA system, or as a part of their publication process each month. These include cross-checks against housing benefit and Fair Rents records, high-low rent checks, removal of duplicates and descriptive errors.

Data collection also undergoes a range of regular auditing, including hard-copy data matching, accompanied visits and follow-up phone calls. Monitoring tools enable collection to be tracked.

There are a range of quality assurance and data validation processes in each collection, which complement those we require and take place either as the data is entered onto the system or as a part of the Local Housing

Allowances (LHA) publication process each month.

## Cross-checks against Housing Benefit (HB) and Fair Rent records

As data is entered onto the rent officer database, the creation or update of an address triggers a cross referencing with HB and Fair Rent records, which allows the rent officer to make a judgement about the nature of the tenancy and exclude lettings fitting the criteria listed above.

## High-low rent check

Rent officers determine the values that represent exceptional or outlier rental values within the context of the list of rents in each broad rental market area. These parameters are applied to the data extract that is used for LHA production each month. Records that fall below the low or above the high value in the relevant category are double-checked against the source data for accuracy. Data that is found to be erroneous is amended or deleted.

## Removal of duplicates

This is a 2-part process. Initially, rent officers are able to view a list of possible duplicate entries when entering their data.

Secondly, a further check takes place as part of the LHA production process to remove any remaining records added to the system in error. The list of possible duplicates is produced from the data extract used for LHA production. A data-matching query compares address, date of collection and rental value fields – these are then checked by rent officers, confirmed as correct or deleted from the database. This exercise is conducted on a monthly basis. As the LHA dataset uses a 12-month date range there are a number of duplicate items that relate to the re-let of a property at the same rental value within the 12-month period. Both records are acceptable for inclusion in the list of rents used for LHA purpose.

## Descriptive errors

The following parameters are applied to each month's data extract as a part of the LHA production process; records are checked against source data and either amended, deleted or confirmed as a correct entry.

There are 23 automated lettings information quality assurance checks in total and 2 additional checks against duplicates and data share. They identify data where:

- 0 value for sole bedrooms (except “bedspaces” and “caravan site rents”)
- 0 value for sole living rooms and not a “room” (letting) or studio
- anything non-self-contained that are not categorised as “rooms”
- number of living rooms exceeds the number of bedrooms
- fuel flagged but no ineligible services figure entered
- incomplete postcode
- exceeds rental parameters (above or below extreme observations determined by local rent officers)
- services ineligible for housing benefit deducted from gross rent is greater than 25% of gross rent
- property condition is missing
- 0 value for sole rooms (except “bedspaces” and “caravan site rents”)
- “house” with 1 sole room (sole rooms are the product of sole bedrooms plus sole living rooms)
- whole “house” or “flat” with shared bedroom, living room, kitchen or bathroom
- “studio” with sole rooms less or greater than 1
- “rooms” (used for letting of 2 or more shared rooms in non-self-contained) with 1 sole room
- “room” (letting) with more than 1 sole room
- tenancy start date is before 15 January 1989, that is, regulated tenancy – deleted centrally by Central Support Team (CST)
- dwelling type equals unique
- property type equals hostel
- “bedspace” with sole rooms is greater than 0
- more than 25% variance between achieved and advertised rent (rents negotiated down or forced up by more than 25% based on supply and demand in the market)
- future tenancy start date – deleted centrally by CST
- incorrect local authority picked from drop-down list
- duplicate checks against other lettings information entered based on an exact match of full address, postcode and sole rooms
- duplicates matched against award data share received from local authorities (people entitled to Housing Benefit); this is based on sole rooms, postcode and first 15 characters of address field – deleted centrally by CST

Typically, about 10% of the data entered during the month is subject to further challenge and scrutiny. Data identified by the CST is reported back to the responsible rent officer. Unsatisfactory responses are returned to the rent officer. CST retain, and will action, data where issues are satisfactorily resolved. Around 3.7% of the data on average goes on to be either amended (approximately 0.9%) or deleted (about 2.8%).

CST also run additional periodic checks for potential duplicate data and other quality assurance checks within the previous 12 months in the run up to significant outputs such the annual LHA determinations. CST adapt their approach to address any potential emerging issues and maintain a close feedback loop with operational management and Information and Analysis directorate.

Following an internal audit of the CPIH process, access to the CPIH-related macros and code has been restricted to those in the private rental market team within Information and Analysis directorate.

Information and Analysis directorate request the data extracts directly from Digital group, logging a new request for the time period to query each month. Contingency has also been put in place so the process can be run by several individuals and at different office locations.

In completing the CPIH production run, quality assurance and data checks are included. These pick up errors in the code, errors in the initial datasets and provide comparisons between previous monthly outputs. Any large movements by region and property type are queried with the housing allowance data collection team to quantify and provide reasoning behind these movements. The code and various outputs are vigorously checked by an independent statistician within the team. If all checks are complete and no errors are identified, the resulting output is signed off by the private rental market lead and sent to us as specified in the SLA.

An illustration of this process can be found in [Annex B: Flow Diagrams of private rents quality assurance processes](#); Figure 4: VOA data flow diagram – data collection and Figure 2: VOA data flow diagram – OOH processing which can be found in section 4.5, 'Areas for improvement'.

## 4.4 Producers' quality assurance investigations and documentation

We have worked closely with the Valuation Office Agency (VOA) to improve the processing, methods and metrics produced as part of the VOA data delivery by updating and quality assuring the SAS code used to produce the data and aggregates as required.

The following 4 areas were identified for methodological improvement and were implemented in March 2015:

- improvements to the process for determining comparable replacement properties when a price update for a sampled property becomes unavailable, leading to more viable matches
- bringing the process for replacing properties for which there is no comparable replacement into line with that used for other goods and services in consumer price statistics
- optimising the sample of properties used at the start of the year, to increase the pool of properties from which comparable replacements can be selected
- reassessing the length of time for which a rent price can be considered valid before a replacement property is found

Further information on these improvements implemented can be found within the published article [Improvements to the measurement of owner occupiers' housing costs and private housing rental prices](#).

Under the SLA, we require 3 months' notice for any changes in processing code, data collection or format of data delivery to ensure sufficient time for amendments and testing. Any changes to the processing code are externally quality assured and signed off by us.

## Data received

Each month we receive several datasets as part of the agreed delivery from VOA:

- elementary aggregates – used in constructing the indices
- diagnostics – used as quality indicators for the process and data
- low-level aggregates – used to investigate movements in the data

Diagnostics and low-level aggregates were added as requirements under the latest SLA and have been received as part of the monthly data delivery since March 2016.

## Diagnostics

We check metrics detailing changes to the sample and stratification levels to ensure the sample remains within acceptable parameters to produce high-quality statistics. A number of metrics are provided, which include information on:

- sample size
- number of replacements required
- number of successful replacements

Additional metrics are derived from this and monitored on a monthly basis:

- reduction in sample size – if there is any drop in sample size within the year the data provider is contacted to clarify the reason for this, as this could indicate an error in the process or data
- replacement success – if the replacement success rate falls below 70% then the data provider is contacted as this could indicate insufficient records in the replacement pool, perhaps caused by changes in collection practices
- percentage of updates – if the percentage of updates falls below 2% then the data provider is contacted, as this could indicate changes in practices in following up properties

An illustration of some of these metrics is provided within [Appendix B of the article Improvements to the measurement of owner occupiers' housing costs and private housing rental prices](#). These thresholds will be reviewed annually as a longer diagnostics series becomes available.

## Elementary aggregates

Elementary aggregate data for England (VOA) is combined with that of Scotland and Wales within our system, with this process being run by 2 individuals independently in what is referred to as a "double run". Any internal processing errors are captured and resolved through this approach.

Month-on-month growth in the index is analysed at region by property type with any movements greater than or less than 1% flagged for further analysis. These can then be queried with the data provider.

The data is then aggregated with the resulting series being analysed at a regional level and checks made between the various measures which are based on the same underlying data – that is, comparisons are made between owner occupiers' housing (OOH), [Index of Private Housing Rental Prices \(IPHRP\)](#), Consumer Prices Index (CPI) rent and Retail Price Index (RPI) rent. Any unexpected movements within the series, which are driven by the raw data, are queried with the data providers who liaise directly with rental officers. Low-level aggregates (sample averages at local authority district level) for England and record-level data for Wales and Scotland are used to explore movements in the index further.

Monthly meetings between operational contacts, noted in 3.2: Communication with data supplier partners, are used to review the latest month and discuss any long-term trends in the data and its drivers.

## Quality management

From September 2016, the processing of OOH costs and private rental data (within the organisation) became part of this quality management system. We commissioned an external audit of VOA rental processes in September 2016 and will continue to do so on an annual basis. The 2016 audit found that “the quality elements of the VOA seen through documents and processes provides confidence that it is able to meet the requirements of the ONS”. This audit followed the previous one, which we commissioned in 2015.

Within our organisation, for items collected as part of the [CPI](#), quality management systems comply with ISO 9001 accreditation and the division is externally audited by [Certification International](#) to ensure the processes and practices are operating effectively and there is continued compliance with the standard.

## Comparisons with other sources

Comparisons are made between the annual growth of IPHRP and that of some private rental providers. This comparison has been published as part of the [IPHRP published tables](#) since the IPHRP April 2016 release. Particular focus is given to comparisons against the occupied lets series published by [Countrywide](#), a stock measure and hence more comparable with IPHRP.

In September 2015, we [published analysis](#) focusing on the difference between our private rental indices and the VOA's private rental market statistics, both of which are based on the same underlying data collected by VOA rent officers. We will look to extend this analysis during 2017. [An article comparing measures of private rental growth was published in October 2016.](#)

Further information on the methods applied and quality checks implemented can be found within the article [Improvements to the measurement of owner occupiers' housing costs and private housing rental prices](#), the [Index of Private Housing Rental Prices Quality and Methodology Information](#) (which uses the same underlying data) and the [CPIH compendium](#).

For a visual summary of Quality Assurance processes undertaken for VOA administrative data, please refer to [Annex B: Flow diagrams of private rents quality assurance processes](#). Figure 1 shows the data collection processes, whilst figure 2 shows the processing that is completed on the data for the OOH component of CPIH.

## 4.5 Areas for improvement

Legislation governing the VOA under the Commissioners of Revenue and Customs Act 2005 means that we do not have access to VOA rental microdata. As such, we have worked closely with colleagues in the VOA to develop their processes and methodology, who then deliver elementary aggregates to us. Access to VOA rental data may be made possible through legislation put forward as part of the [Digital Economy Bill](#).

## 5 . Welsh government

### 5.1 Operational context and admin data collection

Rent Officers Wales is part of the Housing Policy Division of the Welsh government and provides rental data, which is used to construct the Wales estimate. Residential accommodations in the private rented sector in Wales are valued by rent officers who provide an independent and impartial valuation service of residential properties. The market rental evidence team of Rent Officers Wales are in regular contact with landlords and letting agents who provide them with the latest up-to-date information, on a voluntary basis, to ensure all valuations are based on current open market rents.

There are currently 5 Rent Officers in the Market Rental Evidence Team and 1 line manager responsible for data collection within the 22 Broad Rental Market Areas (BRMAs) throughout Wales. There are 7 main collection methods used by Rent Officers:

- visits
- letting lists
- telephone
- mail merge
- via email
- websites
- forums and surveys

Each source is encouraged to provide details of their whole portfolio and to update the data on a regular basis. The information is captured electronically in the Rent Officers Wales lettings information database. Checks are carried out at the point of entry to ensure that any Housing Benefit-funded tenancies are identified.

Rent Officers Wales aim to collect between 15 and 20% of the private rental market across Wales as a whole, excluding lettings known to be subject to Housing Benefit and those with incomplete information. There is no definitive data giving the size or composition of the PRS. The most accurate data currently available is the 2011 Census, so this is taken as the baseline for establishing the required sample.

Data collection is monitored against the private rental market identified by the 2011 Census in an attempt to ensure that the sample is as representative of the market as possible. However, it is dependent upon the goodwill of agents and landlords for its provision. Landlords who only let 1 or 2 properties are contacted once or twice a year to obtain details, whereas agents and those landlords that have large portfolios are contacted frequently for new additions or changes to their letting portfolio. Rent officers also monitor websites and follow up contacts with agents to obtain details as properties are let or removed from the sites.

### 5.2 Communication with data supplier partners

Service Level Agreements (SLAs) between ourselves and the Welsh government document activities such as data requirements, data transfer process and data protection. A delivery schedule is included within an appendix of this SLA which is reviewed annually. This schedule has always been met. Under this SLA, we require 3 months' notice of any changes in data collection practices or format and coding of the data delivery to ensure sufficient time for amendments and testing.

Agreements are reviewed annually to see if the aims, benefits and terms of the SLA are mutually acceptable or require some adjustments. Annual meetings are held to discuss performance and future plans.

As well as these formal arrangements there is also email communication on a monthly basis as part of the quality assurance of the raw price quotes.

### **5.3 Quality assurance principles, standards and checks by data suppliers**

Rent Officers Wales aim to collect between 15 and 20% of the private rental market across Wales as a whole, excluding lettings known to be subject to housing benefit and those with incomplete information. This equates to approximately 2,500 dwellings sampled each month. There is no definitive data giving the size or composition of the complete market. The most accurate data currently available is the 2011 Census so this is taken as the baseline for establishing the required sample.

#### **Quality checks required for Local Housing Allowances (LHA)**

There are a range of quality assurance and data validation processes in each collection, which complement those we require and take place either as the data is entered onto the system or as a part of the LHA publication process each month.

## 1. Cross-checks against Housing Benefit and Fair Rent records

As data is entered onto the rent officer database, the creation or update of an address triggers a cross-referencing with Housing Benefit and Fair Rent records, which allows the rent officer to make a judgement about the nature of the tenancy and exclude lettings fitting the criteria listed at the beginning of this section.

## 2. High-low rent check

Rent officers determine the values that represent exceptional or outlier rental values within the context of the list of rents in each BRMA. These parameters are applied to the data extract that is used for LHA production each month. Records that fall below the low or above the high value in the relevant category are double-checked against the source data for accuracy. Data that is found to be erroneous is amended or deleted.

## 3. Removal of duplicate

This is a 2-part process; initially rent officers are able to view a list of possible duplicate entries when entering their data.

Secondly, a further check takes place as part of the LHA production process to remove any remaining records added to the system in error. The list of possible duplicates is produced from the data extract used for LHA production. A data matching query compares address, date of collection and rental value fields; these are then checked by rent officers, confirmed as correct or deleted from the database. This exercise is conducted on a monthly basis. As the LHA dataset uses a 12-month date range there are a number of duplicate items that relate to the re-let of a property at the same rental value within the 12-month period. Both records are acceptable for inclusion in the list of rents used for LHA purposes.

## 4. Descriptive errors

The following parameters are applied to each month's data extract as a part of the LHA production process; records are checked against source data and either amended, deleted or confirmed as a correct entry:

- zero bedrooms
- non-self-contained properties that are not "rooms"
- number of living rooms exceeds number of bedrooms
- property type "studio" with more than 1 room
- self-contained "room" lettings
- number of bedrooms exceeds 7
- rents listed as including gas and or electricity costs but without a value deducted from the rent for this component

## Audit

### 1. Hard copy data matching

Every month, 10 lettings research rent officers have all the data they have entered during the previous month audited by their manager. This involves each record being checked against the relevant paper-based data sheet that was completed in the field. This process identifies error, omission, accuracy of filing, as well as being a chance to talk to rent officers about collection habits. Each manager completes a short report which summarises findings with any areas for improvement discussed with the rent officer in question.

### 2. Accompanied visits

During the course of the year, each manager spends at least half a day accompanying their rent officers on visits to sources. This provides the manager with a chance to monitor their staff in the field, pick up areas that require development or indeed record best practice that can be shared with other members of lettings research. A short report is completed and shared with the member of staff in question.

### 3. Follow-up phone calls

Managers are encouraged to make a selection of phone calls to data sources that rent officers visit to check on the quality of interaction with the source and if necessary to double-check the data recorded on the system.

An illustration of this process can be found in Figure 6: Welsh government flow diagram – data collection, which can be found following Practice area 4.

## 5.4 Producers' quality assurance investigations and documentation

The Welsh government provides us with microdata on private rental properties, which are loaded into our rental data repository on a monthly basis. The import process recodes and recalculates some of the imported data. It also checks whether the rents are within the boundaries assigned by the user, whether there are any actual duplicates in the file and whether there are any potential duplicates to query with the suppliers. The following types of records are flagged and provided in the reports:

- import errors: these are records that failed to import; for example, they might not contain rental values, have dates earlier or later than the dates expected in the file or be blank
- internal duplicates: these are records that are duplicated in the file being loaded; they have the same addressID, the same rent and the same property attributes
- external duplicates: these are records that already exist in the repository
- addressID queries: these are several records that have the same address identifier in the file being loaded, however, they have some different attributes; they are potentially duplicates
- attribute queries: these are records that match each other in all attributes, rent and date loaded, but have a different address identifier; they are potentially duplicates
- rent queries: these are records that have rents that are higher or lower than the boundaries assigned by the user; they are potentially incorrect
- change queries: these are records of addresses that already existed in the repository; however, the attributes of the address have changed

Records flagged in these reports are queried with the data supplier who then cross-references them against their own database and advises on the correct treatment. An audit trail of all data imported into the repository is kept.

The Welsh government have 48 hours to respond and advise on the correct treatment of the records sent.

Data is then fed through the monthly processing. The monthly calculation is a fairly complex process, but the process of running it is straightforward. Attention is given to whether there are any errors at the different stages. Within the process, summary tables are produced such as a count of records in and out of the sample by property type, country and furnished or unfurnished status.

## Elementary aggregates

Elementary aggregate data for Wales is combined with that of Scotland and England within our system. This process is run by 2 individuals independently in what is referred to as a “double run”. Any internal processing errors are captured and resolved through this approach.

Month-on-month growth in the index is analysed at region by property type with any movements greater than or less than 1% flagged for further analysis. These can then be queried with the data provider.

The data is then aggregated with the resulting series analysed at a regional level and checks made between the various measures that are based on the same underlying data (that is, owner occupiers' housing (OOH), [Index of Private Housing Rental Prices \(IPHRP\)](#), Consumer Prices Index (CPI) rent and Retail Price Index (RPI) rent. Any unexpected movements within the series are investigated using the raw data. If necessary, these are queried with the data provider who can help by advising on perhaps regional policy changes.

## Comparisons with other sources

Comparisons are monitored against other sources in the same way as VOA data, as described in section 4.

## Quality management

For items collected as part of the [CPI](#), quality management systems comply with ISO 9001 accreditation and the division is audited regularly by [Certification International](#) to ensure our systems are operating effectively and there is continued compliance with the standard. From September 2016, the processing of owner occupiers' housing costs and private rental data became part of this quality management system.

The above describes in detail the Quality Assurance processes undertaken for Welsh Government administrative data. For a visual summary of these processes, please refer to [Annex B: Flow Diagrams of private rents quality assurance processes](#). Figure 3 shows the data collection process.

# 6 . Scottish government

## 6.1 Operational context and admin data collection

Rental data for Scotland is currently provided by Rent Service Scotland (formally known as the Rent Registration Service), which is part of the Communities Analysis Division of the Scottish government. It is responsible for gathering rental prices and analysing local rental markets to provide local authorities with Local Housing Allowances (LHA) figures. This information on the rental market is collected by market rental evidence teams, which are in regular contact with landlords and letting agents. There are currently 5 rent officers in the Market Rental Evidence Team and 1 line manager responsible for data collection. There are several collection methods used by rent officers:

- landlords
- letting agencies
- internet listings
- private adverts
- rental forums

Internet listings are by far the most important with the vast majority of rental data being collected from this source. Given this, data for Scotland are mainly based on advertised rather than achieved rents. Evidence published by [Countrywide](#) gives an average asking to achieved rent in Scotland of 99.7%, suggesting perhaps very little difference between both measures.

It is estimated that private landlords make up around 5% of the sample. Scottish government has strong links with associations such as the Scottish Association of Landlords and attend various landlord forums, which are used to identify and maintain data sources.

## **6.2 Communication with data supplier partners**

Service Level Agreements (SLA) between ourselves and the Scottish government document activities such as data requirements, data transfer process and data protection. A delivery schedule is included within an appendix of this SLA which is reviewed annually. This schedule has always been met. Under this SLA, we require 3 months' notice of any changes in data collection practices or format and coding of the data delivery to ensure sufficient time for amendments and testing.

Agreements are reviewed annually to see if the aims, benefits and terms of the SLA are mutually acceptable or require some adjustments. Annual meetings are held to discuss performance and future plans.

As well as these formal arrangements there is also email communication on a monthly basis as part of the quality assurance of the raw price quotes.

## **6.3 Quality assurance principles, standards and checks by data suppliers**

Rent Service Scotland – For Scotland the target sample size is 10% coverage in all designated areas based on sources such as census results and landlord registration data. They use local evidence to ensure the data is representative (of size and type of property) of each area and use geographic information system mapping to supplement local knowledge. This equates to approximately 2,300 dwellings sampled each month.

### **Quality checks required for Local Housing Allowances (LHA)**

There are a range of quality assurance and data validation processes in each collection, which complement those that we require. These take place either as the data is entered onto the system or as a part of the LHA publication process each month.

## 1. Cross-checks against known Housing Benefit and Fair Rent records

As data is entered onto the rent officer database, the creation or update of an address triggers a cross-referencing with Housing Benefit and Fair Rent records. This allows the rent officer to make a judgement about the nature of the tenancy and exclude lettings fitting the criteria listed at the beginning of this section.

## 2. High-low rent check

Rent officers determine the values that represent exceptional or outlier rental values within the context of the list of rents in each Broad Rental Market Area (BRMA). These parameters are applied to the data extract that is used for LHA production each month. Records that fall below the low or above the high value in the relevant category are doubled-checked against the source data for accuracy. Data that is found to be erroneous is amended or deleted.

## 3. Removal of duplicate

This is a 2-part process; initially rent officers are able to view a list of possible duplicate entries when entering their data.

Secondly, a further check takes place as part of the LHA production process to remove any remaining records added to the system in error. The list of possible duplicates is produced from the data extract used for LHA production. A data matching query compares address, date of collection and rental value fields; these are then checked by rent officers, confirmed as correct or deleted from the database. This exercise is conducted on a monthly basis. As the LHA dataset uses a 12-month date range, there are a number of duplicate items that relate to the re-let of a property at the same rental value within the 12-month period. Both records are acceptable for inclusion in the list of rents used for LHA purposes.

## 4. Descriptive errors

The following parameters are applied to each month's data extract as a part of the LHA production process; records are checked against source data and either amended, deleted or confirmed as a correct entry:

- zero bedrooms
- non-self-contained properties that are not "rooms"
- number of living rooms exceeds number of bedrooms
- property type "studio" with more than 1 room
- self-contained "room" lettings
- number of bedrooms exceeds 7
- rents listed as including gas and or electricity costs but without a value deducted from the rent for this component.

## Audit

## 1. Hard copy data matching

Rent officers are responsible for quality assuring all lettings information (LI) entered onto the Rent Officer Case Administration System (ROCAS). Every month, rent officers quality assure (QA) the information entered on or after the 28th of the previous month up to and including the 27th of the current month. The Rental Lettings Team Leader does a quality check at the end of the month before the report is run. The team leader can also perform random sample checking at any time in the ROCAS system.

Any entries made by the rent officers are stored in the Scottish government electronic Record Management (eRDM) either electronically at time of input or scanned in later by the admin team.

ROCAS produces a report, "Local Housing Allowance (LHA) with Letting Information", which the Rental Lettings Team Leader validates with Community Analysis Division (CAD) and the SAS report it produces. We use pivot tables to identify the 30th percentile and a market evidence analysis sheet to identify any drop off data. CAD also produces the 25, 50 and 75 percentiles at a local authority level to help in identifying any gaps in the market. This process identifies error, omission, accuracy of filing as well as being a chance to talk to rent officers about collection habits. The rental team leader completes a short report which summarises findings with any areas for improvement discussed with the rent officer in question.

## 2. Accompanied visits

During the course of the year, the Rental Lettings Team Leader spends at least half a day accompanying their rent officers on visits to sources. This provides the manager with a chance to monitor their staff in the field, pick up areas that require development or indeed record best practice that can be shared with other members of lettings research. A short report is completed and shared with the member of staff in question.

## 3. Follow-up phone calls

Managers are encouraged to make a selection of phone calls to data sources that rent officers visit to check on the quality of interaction with the source and if necessary to double-check the data recorded on the system.

An illustration of this process can be found in [Annex B: Flow Diagrams of private rents quality assurance processes](#); Figure 7 Scottish government flow diagram – data collection, which can be found following Practice area 4.

## 6.4 Producers' quality assurance investigations and documentation

The Scottish government provide us with microdata on private rental properties, which are loaded into our rental data repository on a monthly basis. The import process recodes and recalculates some of the imported data. It also checks whether the rents are within the boundaries assigned by the user, whether there are any actual duplicates in the file and whether there are any potential duplicates to query with the suppliers. The following types of records are flagged and provided in the reports:

- import errors: these are records that failed to import; for example, they might not contain rental values, have dates earlier or later than the dates expected in the file or be blank
- internal duplicates: these are records that are duplicated in the file being loaded; they have the same addressID, the same rent and the same property attributes
- external duplicates: these are records that already exist in the repository
- addressID queries: these are several records that have the same address identifier in the file being loaded, however, they have some different attributes; they are potentially duplicates
- attribute queries: these are records that match each other in all attributes, rent and date loaded, but have a different address identifier; they are potentially duplicates
- rent queries: these are records that have rents that are higher or lower than the boundaries assigned by the user; they are potentially incorrect
- change queries: these are records of addresses that already existed in the repository; however, the attributes of the address have changed

Records flagged in these reports are queried with the data supplier who then cross-references them against their own database and advises on the correct treatment. An audit trail of all data imported into the repository is kept.

Data is then fed through the monthly processing. The monthly calculation is a fairly complex process, but the process of running it is straightforward. Attention is given to whether there are any errors at the different stages. Within the process, summary tables are produced, such as a count of records in and out of the sample by property type, country and furnished or unfurnished status.

## Elementary aggregates

Elementary aggregate data for Scotland is combined with that of Wales and England within our system. This process is run by 2 individuals independently in what is referred to as a "double run". Any internal processing errors are captured and resolved through this approach.

Month-on-month growth in the index is analysed at region by property type, with any movements greater than or less than 1% flagged for further analysis. These can then be queried with the data provider.

The resulting series is analysed at a regional level and checks made between the various measures that are based on the same underlying data (that is, owner occupiers' housing (OOH), [Index of Private Housing Rental Prices \(IPHRP\)](#), Consumer Prices Index (CPI) rent and Retail Price Index (RPI) rent). Any unexpected movements within the series are investigated using the raw data. If necessary, these are queried with the data provider who can help by advising on perhaps regional policy changes.

## Comparisons with other sources

Comparisons are monitored against other sources in the same way as VOA and Welsh government data.

## Quality management

For items collected as part of the [CPI](#), quality management systems comply with ISO 9001 accreditation and the division is audited regularly by [Certification International](#) to ensure our systems are operating effectively and there is continued compliance with the standard. From September 2016, the processing of owner occupiers' housing costs and private rental data became part of this quality management system.

The above describes in detail the Quality Assurance processes undertaken for Scottish Government administrative data. For a visual summary of these processes, please refer to [Annex B: Flow Diagrams of private rents quality assurance processes](#). Figure 3 shows the data collection process.

# 7. Annex A: New quality assurance practices for private rental data



External quality assurance	<p>We commissioned an external audit of Valuation Office Agency (VOA) rental processes in September 2016. This follows on from the previous external audit we commissioned in 2015. The 2016 report found that;</p> <p>"Overall, the quality elements of the VOA seen through documents and processes provides confidence that it is able to meet the requirements of the ONS."</p> <p>VOA have agreed that they are happy for us to continue to commission an external audit on an annual basis. The Service Level Agreement (SLA) for 2017 is currently being drafted. The requirement for an external audit will be included as part of this Service Level Agreement.</p>
Improved communication between ONS and VOA	<p>The 2016 SLA which came into effect for the February 2016 delivery in March 2016 provided further clarification and structure for the communication between us and VOA regarding the processing of private rental data.</p> <p>Regular meetings between us and VOA are timetabled to discuss performance and future plans. These include monthly, quarterly and half-yearly meetings between operational, ONS and information asset owner contacts.</p>
Quality standards of VOA	<p>Our quality standards and quality assurance mechanisms were implemented to VOA processing when re-developing the VOA code, which was subsequently quality assured by the methodology division within ONS. Under the SLA if VOA are to make any changes to its processes they are required to provide us with 3 month's prior notice so that ONS may quality assure and sign off this change.</p>
Potential sources of bias and error	<p>Potential sources of biases and implications on the quality of CPIH are discussed within the CPIH compendium. Potential sources of error in VOA administrative data and how VOA look to minimise these through quality assurance are included within the VOA assessment.</p>
Monitoring of metrics	<p>Metrics have been formally monitored since March 2016 (reference period Feb 2016). These metrics were received on an ad hoc basis prior to March 2016 but now formally form part of the 2016 SLA and delivery schedule.</p>
Comparison of OOH and IPHRP with VOA statistics	<p>An <a href="#">update to this article</a> was published in October 2016, which looked to extend the current time series published. Additional resources which have been allocated to this task can be found in our <a href="#">development plan</a>.</p>
ISO 9001 Accreditation	<p>Audited and accredited in September 2016. Now forms part of Prices regular audit.</p>
Welsh and Scottish government SLA	<p>Welsh and Scottish government quality assurance mechanisms are documented within the QAAD. SLAs are updated on an annual basis and have an agreement that any queries are answered within 48 hours, which has always been met. This will be included within the 2017 SLA, which is currently being drafted.</p>
Explanation of potential bias in Scotland	<p>The potential bias relating to the fact that rent officers in Scotland collected advertised rather than achieved rent prices is now explained within the Scottish government assessment, referencing evidence by countrywide.</p>
Process map of Welsh and Scottish data collection	<p>This is now included as part of the Welsh and Scottish government assessments.</p>

Information management and governance groups	A chart describing the role of relevant information management and governance groups in data quality management is included in <a href="#">Annex B</a> .
--	--

## 8. Annex B: Flow diagrams of private rents quality assurance processes

[Annex B: Flow diagrams of private rents quality assurance processes](#)