

Quality assurance of administrative data used in consumer price inflation statistics

This publication documents the quality assurance (QA) assessment of all data sources used in the production of the Consumer Price Index with Owner Occupiers' Housing Costs (CPIH).

Contact:
Chris Payne
chris.payne@ons.gsi.gov.uk
+44 (0)1633 455321

Release date:
19 July 2017

Next release:
To be announced

Table of contents

1. [Executive summary](#)
2. [Introduction](#)
3. [Quality considerations](#)
4. [Approach to assessment](#)
5. [Assurance level assessment](#)
6. [Action plan](#)
7. [Annex A: Assessment of data sources](#)
8. [Annex B: ONS data checking and validation](#)
9. [Annex C: Summary of monthly validation](#)
10. [Annex D: Quality assurance processes and checks – UK Consumer Research](#)
11. [Annex E: Flow diagrams of quality assurance processes](#)

1 . Executive summary

There are a range of consumer price inflation measures in use in the UK; notably the Consumer Prices Index including owner occupiers' housing costs (CPIH), and the Consumer Prices Index (CPI), which omits these housing costs.

CPIH is the first measure of inflation in our Consumer Price statistics bulletin. It was launched in 2013, but was subsequently de-designated as a National Statistic following the identification of required improvements to the methodology. We have now implemented all of the improvements, and are seeking re-designation for CPIH as a National Statistic.

The construction of CPIH and CPI is complex. Price and expenditure data are required for each of the approximately 700 items in the "basket" of goods and services. A variety of different data sources are used for this purpose.

The data used in the compilation of CPIH and CPI can be categorised as follows:

1. Price collection from shops in various locations around the country (commonly referred to as the "local" collection), which is contracted to an external company called TNS
2. Individual prices collected through a website, phone call to the supplier, or from a brochure.
3. Expenditure weights or prices calculated from survey data, which are sourced from within ONS, or from another government department.
4. Expenditure weights or prices calculated from administrative data, which taken from or compiled within ONS, other government departments, or commercial companies.

The owner occupiers' housing costs (OOH) component of CPIH uses 4 administrative data sources to calculate the cost of owning, maintaining and living in one's home; these are sourced from the Valuation Office Agency (VOA) in England, and from the Welsh and Scottish governments (Northern Ireland data are currently used from the TNS collection). Data from the Department for Communities and Local Government (DCLG) are used to weight price data to reflect the owner occupied housing market.

Incorporating so many different data sources into any statistic, but particularly one used as a key economic measure, involves a certain degree of risk. Administrative data in particular may be collected and compiled by third parties, outside the Code of Practice for official statistics.

Our production processes are certified under an external quality management system: ISO9001: 2015. However, to further assure ourselves and users of the quality of our statistics, we have undertaken a thorough quality assessment of these data sources. This assessment is a continuous process, and we will publish updates periodically.

We have followed the Quality Assurance of Administrative Data (QAAD) toolkit, as described by the Office for Statistics Regulation (OSR). Using the toolkit, we established the level of assurance we are seeking (or "benchmark") for each source. The assurance levels are set as either "basic", "enhanced" or "comprehensive", depending on:

- the risk of quality concerns for that source, based on various factors, such as the source's weight in the headline index, the complexity of the data source, contractual and communication arrangements currently in place, and other important considerations
- the public interest profile of the item which is being measured, and its contribution to the headline index

The majority of items in the consumer prices basket of goods and services are constructed from just three key sources of data: the local price collection from TNS, expenditure data from Household Final Consumption Expenditure in the national accounts, and further expenditure data from the Living costs and Food Survey. This means that there are a few sources which will need a higher level of assurance, and many sources which are only used for one component of the index and so do not require a particularly high level of assurance.

Through engagement with our suppliers, we have assessed the assurance level that we have currently achieved by considering:

- the operational context of the data; why and how it is collected
- the communication and agreements in place between ourselves and the supplier
- the quality assurance procedures undertaken by the supplier
- the quality assurance procedures undertaken by us

Table 1 below summarises the quality assurance benchmarks that were set, and the assurance levels that we have assessed each source at during this assessment.

Table 1: Summary of Benchmark and actual assessment levels for CPIH data sources

	Risk	Profile	Benchmark QA level	Achieved QA assessment
VOA	High	High	Comprehensive	Enhanced
Welsh Government	Low	Medium	Enhanced	Comprehensive
Scottish Government	Low	Medium	Enhanced	Comprehensive
Mintel	Medium	Medium	Enhanced	Comprehensive
Glasses	Low / Medium	Low	Basic	Basic (Incomplete)
Moneyfacts	Low	Low	Basic	Basic
HESA	Low	Low	Basic	Basic
Consumer Intelligence	Low	Low	Basic	Basic
Kantar	Low	Low	Basic	Basic
IDBR	Low	Low	Basic	Basic
Website	Low	Low	Basic	Basic
Direct Contact	Low	Low	Basic	Basic
Brochures	Low	Low	Basic	Basic
HHFCE	High	High	Comprehensive	Enhanced
LCF	Low	High	Enhanced	Comprehensive
TNS	Medium	High	Comprehensive	Comprehensive
DCLG	Medium	High	Comprehensive	Enhanced
BEIS	Low	Low	Basic	Basic/Enhanced
IPS	Low	Low	Basic	Basic/Enhanced
Home and Communities agency	Low	Low	Basic	In Progress
Department for Transport	Low	Low	Basic	In Progress

Source: Office for National Statistics

As a result of this assessment, we have put in place an action plan to improve our quality assurance in some areas:

- Data for the Valuation Office Agency (VOA) require a comprehensive level of assurance; however, we do not currently have access to the microdata, which limits our ability to quality assure the data; we have mitigated the risks involved by working with VOA to put in place aggregation methodologies, processes and quality assurance procedures, and will request direct access once the Digital Economy Bill has been passed
- Household Final Consumption Expenditure (HHFCE) data also require a comprehensive level of assurance; however, we would like more information on the complex array of data sources used to compile the statistics. HHFCE are developing a QAAD assessment, and expect to deliver this in autumn 2017
- similarly DCLG data require a comprehensive level of assurance; however, the data are constructed from a number of data sources, which have not necessarily been comprehensively quality assured. We are working with DCLG to understand their departmental approach to the development of QAAD material
- finally, there are a number of data sources which require basic assurance, for which we have not received all the requested quality assurance information; we will work with these suppliers to gain the level of assurance we require

We will continue to engage with our data suppliers to better understand any quality concerns that may arise, and to raise their understanding of how their data are used in the construction of consumer price inflation measures. We aim to publish an update to this QAAD in summer 2017.

2 . Introduction

There are currently two key consumer price inflation measures in the UK. The Consumer Prices Index including owner occupiers' housing costs (CPIH) is the first measure of consumer price inflation in our statistical bulletin, and is currently the most comprehensive measure of inflation. This addresses some of the shortcomings of the Consumer Prices Index (CPI), which is an internationally comparable measure of inflation, but does not include a measure of owner occupiers' housing costs (OOH): a major component of household budgets¹. Both of these measures are based on the same data sources (with the exception of OOH and Council Tax, which are in CPIH but not CPI). These data sources are numerous and often complex. We therefore seek to assess the quality of each of these sources.

Our assessment of data sources is carried out in accordance with the [Office for Statistics Regulation's Quality Assurance of Administrative Data \(QAAD\) toolkit](#). We are striving for a proportionate approach in assessing the required level of quality assurance for the many and varied data sources used in the compilation of CPI and CPIH. We seek to highlight and address the shortcomings that we have identified, and reassure users that the quality of the source data is monitored and fit for purpose.

In this paper, we set out the steps we have taken to quality assure our data, and our assessment of each source. In section 3 we discuss important quality considerations for CPIH and CPI. In section 4 we outline our approach to assessing our data sources. In section 5 we discuss the assurance levels we are seeking for each data source, and the resulting assessment and, in section 7, we detail our next steps towards achieving full assurance. Our detailed quality assurance information for each source is provided in Annex A.

This publication is part of an ongoing process of dialogue with our suppliers, to increase our understanding of any quality concerns in the source data, and to raise awareness of how it is utilised. Through this document, we aim to provide information and assurance to users that the sources used to construct our consumer price inflation measures are sufficient for the purposes for which they are used. We will therefore review this document every 2 years. We do not address the construction of, or rationale for, our OOH measure in CPIH here. This is discussed in detail in the [CPIH Compendium](#). For more information on our consumer price inflation measures, please refer to our [Quality and Methodology Information page](#).

Notes for: Introduction

1. The Retail Prices Index (RPI) is a legacy measure, only to be used for the continuing indexation of index linked gilts and bonds. It is not a National Statistic.

3 . Quality considerations

When considering the quality of UK consumer price inflation measures, there are some broader considerations that users should bear in mind. The first is the de-designation of CPIH as a National Statistic in 2014. The second is external accreditation under ISO9001:2015 for consumer price statistics processes. These are described in more detail in this section. Detail on the quality assurance procedures applied to our statistics is reproduced in Annex B.

3.1 Loss of National Statistics status

CPIH was introduced in early 2013, following a lengthy development process overseen by the Consumer Prices Advisory Committee (CPAC) between 2009 and 2012. CPIH became a National Statistic in mid-2013, but was later de-designated in 2014 after required improvements to the OOH methodology were identified. These were:

- 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

The required methodological improvements were implemented in 2015, and the series was fully revised to accommodate these changes. On 3 March 2016, the Office for Statistics Regulation (OSR) released their assessment report on CPIH, reviewing the statistic against all areas of the Code of Practice for Official Statistics.

We have subsequently undertaken an assessment of all data sources used in the production of CPIH using the OSR's Quality Assurance of Administrative Data toolkit (QAAD). We have aimed to demonstrate that we have investigated, managed and communicated appropriate and sufficient quality assurance of all our data sources. Additionally, we have published a range of supporting information, such as the [CPIH Compendium](#), which sets out the rationale for our choice of OOH measure, and the methodology behind it, the [Comparing measures of private rental growth in the UK](#) article, and the [Understanding the different approaches of measuring owner occupiers' housing costs](#) article. We continue to prioritise work leading to the re-designation of CPIH as a National Statistic.

3.2 ISO9001 Accreditation

Prices Production areas are externally accredited under the quality standard [ISO9001:2015](#). This is an international standard based on a set of quality management principles:

- customer focus
- leadership
- engagement of people
- process approach
- improvement
- evidence-based decision making
- relationship management

It promotes the adoption of a process approach, which will enable understanding and consistency in meeting requirements, considering processes in terms of added value, effective process performance and improvements to processes based on evidence and information. In other words, the main purpose of this standard is to ensure the quality of our production processes, to ensure that we fully evaluate risks and to ensure that we strive for continuous improvement.

The standard is applied to all areas of production involved in the compilation of the whole range of consumer price inflation statistics. Prices documentation is reviewed by trained internal auditors, based on an annual cycle planned by the quality manager. The depth of the audit is based on how frequently the processes change. A review by an external auditor is also conducted on an annual basis, and a 3-year strategic review is also conducted to assess suitability for re-certification.

4 . Approach to assessment

We have conducted our assessment of data sources used in Consumer Prices Index including owner occupiers' housing costs (CPIH) using the Office for Statistics Regulation's QAAD toolkit. We took the following steps for each data source:

- establish the risk of quality concerns with the data
- establish the level of public interest in the item that the data are being used to measure
- determine benchmark quality assurance levels, based on the risk and public interest.
- contact the suppliers of administrative data to understand their own practices and approach to quality assurance; generally, this consists of the following steps:
 - send out questionnaires to our data suppliers requesting information on their QA procedures
 - conduct follow up meetings with our data suppliers to request further information and clarification
 - maintain ongoing dialogues with data suppliers to develop a better understanding of any quality issues in the data, and raise awareness of how the source data are used
- review our own quality assurance and validation procedures and processes
- conduct an assessment of each data source using the four practice areas of the Quality Assurance of Administrative Data (QAAD) toolkit:
 - operational context and data collection
 - communication with data suppliers
 - quality assurance procedures of the data supplier
 - quality assurance procedures of producer
- determine an overall quality assurance level based on our assessment
- if this assurance level does not match the benchmark assurance level, then put steps in place to work towards meeting the required assurance level
- review the quality assurance on an ongoing basis; we will publish a QAAD update every 2 years

4.1 Setting the benchmarks

In accordance with the QAAD toolkit, we have sought assurance for each data source based on the risk of quality concerns associated with that data source, and the public interest in the particular item being measured by that data source.

We considered a high, medium or low risk of data quality concerns based on:

- the weight that the item being measured by a particular data source carries in headline CPIH or Consumer Price Index (CPI); we consider items with a weight less than 1.5% to be very small, items with a weight between 1.5% and 5% to be small, items with a weight between 5% and 10% to be medium, and items with a weight higher than 10% to be large.
- the complexity of the data source; for example, whether it is compiled from a number of different sources, or based on survey data, which we would consider to be lower risk due to the fact that data are collected for statistical purposes and have a holistic, well designed collection strategy, their reliability is better understood, and quality assurance and validation procedures are typically robust
- the existing contractual and communication arrangements currently in place
- how much the measurement of a particular item depends on that data source (in other words, what would we do if we did not have this data?)
- other considerations, such as any existing published information on data collection, methodology or quality assurance, or mitigation of high risk factors with the data

We considered a high, medium or low public interest profile based on:

- the level of media or user interest in the particular item being measured
- the economic or political importance of the particular item being measured
- the contribution of the item being measured to the headline index, since we would consider both CPIH and CPI to be economically and politically important
- any additional scrutiny from commentators, based on particular concerns about the data

Together the risk of quality concerns and public interest profile are combined to set an overall assurance level that is required for a particular source. This assessment is based on the following matrix, as provided by UK Statistics Authority (Table 2).

Table 2: Quality assurance matrix

Level of risk of quality concerns	Public interest profile		
	Lower	Medium	Higher
Low	Statistics of lower quality concern and lower public interest [A1]	Statistics of low quality concern and medium public interest [A1/A2]	Statistics of low quality concern and higher public interest [A1/A2]
Medium	Statistics of medium quality concern and lower public interest [A1/A2]	Statistics of medium quality concern and medium public interest [A2]	Statistics of medium quality concern and higher public interest [A2/A3]
High	Statistics of higher quality concern and lower public interest [A1/A2/A3]	Statistics of higher quality concern and medium public interest [A3]	Statistics of higher quality concern and higher public interest [A3]

Source: UK Statistics Authority

4.2 QAAD practice areas

We have aimed to assess the quality of each data source based on four broad practice areas. These relate to the quality assurance of official statistics and the administrative data used to produce them: our knowledge of the operational context in which the data are recorded, building good communication links with our data suppliers, an understanding of our suppliers' quality processes and standards, and the quality processes and standards that we apply. This is in line with the Office for Statistics Regulations expectations for quality assurance of data sources. The full assessments for each data source can be found in Annex A. Table 3 provides a breakdown of these practice areas.

Table 3: Four practice areas associated with data quality

Operational context and admin data collection	Communication with data partners	QA principles, standards and checks by data suppliers	Producers' QA investigations and documentation
<ul style="list-style-type: none"> • environment and processes for compiling the administrative data 	<ul style="list-style-type: none"> • collaborative relationships with data collectors, suppliers, IT specialists, policy and operational officials 	<ul style="list-style-type: none"> • data assurance arrangements in data collection and supply 	<ul style="list-style-type: none"> • QA checks carried out by statistics producer
<ul style="list-style-type: none"> • factors which affect data quality and cause bias 	<ul style="list-style-type: none"> • formal agreements detailing arrangements 	<ul style="list-style-type: none"> • quality information about the data from suppliers 	<ul style="list-style-type: none"> • quality indicators for input data and output statistics
<ul style="list-style-type: none"> • safeguards which minimise the risks 	<ul style="list-style-type: none"> • regular engagement with collectors, suppliers and users 	<ul style="list-style-type: none"> • role of operational inspection and internal /external audit in data assurance processes 	<ul style="list-style-type: none"> • strengths and limitations of the data in relation to use
<ul style="list-style-type: none"> • role of performance measurements and targets; potential for distortive effects 			<ul style="list-style-type: none"> • explanation for users about the data quality and impact on the statistics

Source: UK Statistics Authority

5 . Assurance level assessment

5.1 Setting the benchmarks

In this section we describe each of our data sources, and consider the assurance level that we are seeking (or "benchmark") for these. We also summarise our current assessment of the data and outline any further steps that may be required to reach the benchmark assurance level. We will also use this process to build engagement with our suppliers to better understand the data source, as well as raising awareness of how the data are used in consumer price inflation statistics.

In the section that follows, the weights provided are for Consumer Prices Index including owner occupiers' housing costs (CPIH) (the first measure of consumer price inflation in our bulletin) in February 2017. These can be considered relatively similar between years. Consumer Prices Index (CPI) weights will be a little higher (as the exclusion of Council Tax and owner occupiers' housing costs will increase the relative share of other items); however, the analysis remains the same.

It is a feature of consumer price statistics that we require a data source for each of the approximately 700 items in the basket of goods and services. The majority of the index is constructed from just three data sources – the local price collection, conducted by an external company called TNS, and expenditure data from the Living Costs and Food Survey (LCF) and the Household Final Consumption Expenditure (HHFCE) branch of the national accounts.

Remaining items tend to be constructed from data sources which are quite specific to the item being measured. A consequence of this is that the distribution of assurance levels required for assessment is very heavily weighted towards basic assurance. This is because we have a few data sources which are used for the vast majority of items, and relatively few items which all require a bespoke data source.

Benchmark assurance levels are summarised in Table 4. The assurance levels required for this QAAD assessment are set out in detail below, with explanations provided accordingly. The assurance levels are based on an assessment of the risk of quality concerns, and the public interest profile, as described in section 3.2. These are used to set the overall assurance level.

Table 4: Benchmark assurance levels and assessment

--

	Risk	Profile	Overall		
	Benchmark Risk Assessment	Benchmark Risk Assessment	Benchmark Profile Assessment	Justification	Actual Risk Assessment
DCLG	Medium	High	Comprehensive	<p>1) Some complexity in data due to various different sources being used in compilation.</p> <p>2) Little information on what QA procedures are applied to these sources by supplier.</p> <p>3) As part of OOH component, this is economically important, albeit with less impact than VOA.</p>	Enhanced: Not Achieved
HHFCE	High	High	Comprehensive	<p>1) A complex data source compiled from numerous different data sources.</p> <p>2) HHFCE are extensively used in CPIH, with no alternative data available.</p> <p>3) Regular communication with supplier and some information on methodology is published.</p>	Enhanced: Not Achieved
TNS	Medium	High	Comprehensive	<p>1) Data accounts for high proportion of prices in CPIH and CPI.</p> <p>2) A dedicated contract management branch assesses TNS's performance against contract.</p> <p>3) Sampling frame and design set by Prices, with quality checks carried out by both parties.</p>	Comprehensive: Achieved

Value Office Agency	High	High	Comprehensive	<p>1) A relatively high weight in CPIH.</p> <p>2) No access to microdata, only aggregated indices, thus limiting Quality Assurance.</p> <p>3) OOH costs economically important, with high user interest in methodology.</p>	Enhanced: Not Achieved
LCF	Low	High	Enhanced	<p>1) Survey data which represents most items at lower level aggregation.</p> <p>2) Collection, design and methodology produced by ONS and is well documented.</p> <p>3) Data are used widely in construction of economically important CPIH and CPI.</p>	Enhanced: Achieved
Mintel	Medium	Medium	Enhanced	<p>1) Although data collection is complex, the process and procedures are well documented.</p> <p>2) A contract is in place with a designated contact.</p> <p>3) Do not represent as broad a cross-section of the basket as other data sources.</p>	Comprehensive: Achieved
Scottish government	Low	Medium	Enhanced	<p>1) Very low weight component of CPIH.</p> <p>2) Unlike VOA, microdata provided, allowing thorough quality assurance.</p> <p>3) Less user and media interest on devolved regions.</p>	Enhanced to Comprehensive: Achieved

Welsh government	Low	Medium	Enhanced	<p>1) Very low weight component of CPIH.</p> <p>2) Unlike VOA, microdata provided, allowing thorough quality assurance.</p> <p>3) Less user and media interest on devolved regions.</p>	Enhanced to Comprehensive: Achieved
BEIS	Low	Low	Basic	<p>1) Data sources are collected through a survey, with relatively low weight contribution.</p> <p>2) Methodology and Quality assurance procedures are considered sufficient, and there is a dedicated BEIS contact.</p> <p>3) Limited, niche media interest in items.</p>	Basic to Enhanced: Achieved.
Brochures	Low	Low	Basic	<p>1) Small to medium weight, with number of mitigating factors that reduce the risk, for instance being collected in-house.</p> <p>2) Manually entered into prices system with robust quality assurance and validation.</p> <p>3) Little media interest in items.</p>	Basic: Achieved
Consumer Intelligence	Low	Low	Basic	<p>1) Very low weight in CPIH, with clear contingency should data become unavailable.</p> <p>2) Contract in place with regular supplier meetings.</p> <p>3) Little media interest in item index.</p>	Basic: Achieved

Department for Transport	Low	Low	Basic	<p>1) Contributes a small to medium weight, with suitable alternatives if data unavailable.</p> <p>2) Data sourced through email contact and imported directly into prices systems.</p> <p>3) Series is of limited media and user interest.</p>	In Progress
Direct Contact	Low	Low	Basic	<p>1) Small to medium weight, with number of mitigating factors that reduce the risk, for instance being collected in-house.</p> <p>2) Manually entered into prices system with robust quality assurance and validation.</p> <p>3) Little media interest in items.</p>	Basic: Achieved
Glasses	Low to Medium	Low	Basic	<p>1) Complex data source compiled from several different sources.</p> <p>2) Relatively small weight contribution and alternative data sources available.</p> <p>3) Detail of quality assurance not yet provided.</p>	Basic: Incomplete
HESA	Low	Low	Basic	<p>1) Very low weight in CPIH, with clear contingency should data become unavailable.</p> <p>2) Data sourced through email with no contractual agreement in place.</p> <p>3) Little media interest in item index.</p>	Basic: Achieved

Home and Communities agency	Low	Low	Basic	<p>1) Very low weight in CPIH, with clear contingency should data become unavailable.</p> <p>2) Data sourced through email with no contractual agreement in place.</p> <p>3) Little media interest in item index.</p>	Basic: Achieved
IDBR	Low	Low	Basic	<p>1) Very low weight in CPIH, with clear contingency should data become unavailable.</p> <p>2) IDBR team is based within ONS, so no contract is in place.</p> <p>3) Little media interest in item index.</p>	Basic: Achieved
IPS	Low	Low	Basic	<p>1) Low but not insignificant weight in headline CPIH.</p> <p>2) Straightforward collection; primarily survey, supplemented by administrative.</p> <p>3) Methodology and Quality Assurance are well documented.</p>	Basic to Enhanced: Achieved
Kantar	Low	Low	Basic	<p>1) Very low weight in CPIH, with clear contingency should data become unavailable.</p> <p>2) Data purchased annually with a dedicated contact provided.</p> <p>3) Little media interest in item index.</p>	Basic: Achieved

Moneyfacts	Low	Low	Basic	1) Very low weight in CPIH, with clear contingency should data become unavailable. 2) Data acquired through an annual magazine subscription. 3) Little media interest in item index.	Basic: Achieved
Website	Low	Low	Basic	1) Small to medium weight, with number of mitigating factors that reduce the risk, for instance being collected in-house. 2) Manually entered into prices system with robust quality assurance and validation. 3) Little media interest in items.	Basic: Achieved

Source: Office for National Statistics

5.2 Assurance level: Comprehensive

We have assessed four of our data sources as requiring a comprehensive level of assurance. This means that we require a detailed understanding of the operational context in which data are collected, including sources of bias, error and mis-measurement. We also require strong collaborative working relationships with these suppliers, supported by firm agreements for data supply, and a detailed understanding of the supplier's quality assurance principles and checks. Our own quality assurance and validation checks should be comprehensive and transparent, and we will communicate any risks that arise from the data.

More detail is provided for each of these four suppliers.

Department for Communities and Local Government (DCLG)

Data usage

Department for Communities and Local Government (DCLG) dwelling stock counts are used in the construction of the owner occupiers' housing costs (OOH) index in CPIH. The English regions, which use DCLG data, have a high weight of 15.14%. The data are used in conjunction with average prices to calculate expenditure totals, which are used to reflect the owner occupied market.

Risk: Medium

Although the OOH component has a high weight in CPIH, DCLG are used below the item level as strata weights, to mix adjust (see the [CPIH Compendium](#)) the private rental price indices to reflect the owner occupied population. Whilst strata weights typically have a more limited impact on the headline measure than higher level weights, for the OOH component, strata level weights are used to distinguish the OOH index from the private rental index. DCLG use a few different sources to compile dwelling stock data. Therefore the source has some complexity; however, not on the level of, say, the national accounts. There is no clear alternative source of data.

Large, volatile movements in the data are uncommon, which means that it is typically quite noticeable if there are issues with the data. This provides some assurance over quality concerns. Some information on DCLG quality assurance and validation procedures is available from their website; however, there is little information on what quality assurance procedures are applied to their data sources. There is a designated contact for DCLG, although no Service Level Agreement (SLA) is in place.

Therefore, a medium risk profile will be applied to DCLG data.

Profile: High

As with VOA data, the OOH component is an economically important component of CPIH, and is the focus of much user attention. Therefore a high public interest profile is appropriate.

Assessment: Enhanced (A2)

Status: Not achieved

DCLG have provided detail of their own quality assurance and validation procedures, which we have reproduced in Annex A. We have assessed the checks and processes as being fit for the purpose for which they are used in the production of CPIH. However, DCLG dwelling stock counts are compiled from a number of administrative and survey data sources and, at present, more information is required on the quality of the sources that DCLG use in the calculation of dwelling stock estimates, and their suitability for this purpose.

Since the previous update to our QAAD we have worked with DCLG who have provided us with a statement on the fitness for purpose of each of their data sources (Annex A), as well as some detail on how consistency is ensured in the data returns that they receive (Annex A). This helps to provide some of the reassurances that a fuller DCLG QAAD would. We have also spoken to other users of DCLG dwelling stock data and established that they are also conducting effective quality assurance of the data, providing an additional layer of assurance (see Annex A). Finally, we have investigated the potential impact of errors in the DCLG data on the OOH component of CPIH and found the impact to be negligible. This analysis will be presented in the Spotlight section of the article Understanding the different approaches to measuring owner occupiers' housing costs, to be published in September. We therefore consider that these actions mitigate for the risks associated with not having achieved the required level of quality assurance for this source.

We have a designated contact for supply of the data; however, regular supplier meetings are not held, and there is no Service Level Agreement in place.

Remedial actions:

1. Seek to understand DCLG's departmental approach to QAAD development, and where dwelling stock count data fits into this process
2. Seek to establish a more robust delivery schedule, including annual supplier liaison meetings, and a Service Level Agreement for delivery

Household Final Consumption Expenditure (HHFCE)

Data usage

CPIH and CPI follow the Classification Of Individual Consumption According to Purpose (COICOP). Expenditure for COICOP categories are used to aggregate lower level indices together. Expenditure weights are based entirely on HHFCE data, produced by the national accounts. Data are taken from the Quarter 3 Consumer Trends publication, which is consistent with the latest Blue Book. Expenditure data are price updated to the relevant period, before being rescaled to parts per thousand for use as expenditure weights. For this reason we could consider HHFCE data to have an almost 100% weight in both CPIH and CPI.

Risk: High

HHFCE is a complex data source, compiled from the Living Costs and Food Survey (LCF), and numerous other administrative sources. Adjustments are also applied to the data; for example, for under-reporting and national accounts balancing. HHFCE data are produced within the ONS.

These data have a very high weight in CPIH and CPI, and there is no real alternative to this source. HICP regulations state that these data must be used as the source of weights for CPI. HHFCE data, however, are also required under European legislation and, as a key component of the national accounts, it is unlikely that they would be discontinued. Data are provided to Prices Division in spreadsheet form, which are fed into Prices systems, and Prices staff will comprehensively quality assure the data.

Some information on methodology, and quality assurance processes is [published](#), and Prices Division have a regular communication mechanism with national accounts staff through a quarterly internal stakeholder board.

Considering the complexity of the data source, and the importance of the data to production of CPIH and CPI, we feel that a high-risk profile is appropriate.

Profile: High

Given the extremely wide coverage of HHFCE data, we have expenditure weights for COICOP categories of varying user interest. Moreover, given that CPIH and CPI are economically and politically important, and HHFCE data are used for all classes, it would be inappropriate to consider anything other than a high public interest profile.

Assessment: Enhanced (A2)

Status: Not achieved

HHFCE expenditure data are compiled from a complex range of administrative and survey data. HHFCE have detailed all of the sources used; however, there is not necessarily detailed quality assurance information provided for each of these data sources. HHFCE have provided detailed information on their quality assurance and validation procedures compilation process, coverage, and forecasting and imputation procedures, which we consider to be fit for purpose. These are reproduced in detail in Annex A.

Prices Division communicates regularly with HHFCE staff through the Prices Stakeholder Board, and there is a good awareness of how HHFCE data are used within consumer price statistics. HHFCE follow international standards; in particular, the European System of National Accounts 2010.

Remedial actions:

1. HHFCE are in the process of completing a QAAD assessment of their data sources. They aim to complete their QAAD assessment by autumn 2017.

TNS

Data usage

Prices for approximately 520 of the items in the consumer prices basket of goods and services are collected from stores and venues across the country by a team of “local” price collectors. The collection is currently carried out by TNS. The total weight of items in the basket collected under local price collection could be as much as 40%.

Risk: Medium

Quality assurance for the local price collection is already well established. A contract is already in place to ensure ongoing price collection, and to ensure that the collection meets the required standard, including what data will be provided, when they will be provided by and in what form TNS will provide them. Prices Division has a dedicated contract management branch that assess TNS’s performance against the contract, using pre-established key performance indicators. Performance is reviewed with the supplier on a monthly basis.

The sampling frame and sample design are specified by Prices Division, and quality checks are carried out on the data by both Prices staff and TNS staff. The quality checks are transparent and clear on both sides, and the process for compiling the data is well established, well documented, and accredited under ISO9001: 2015 by an external body.

TNS data account for a very high proportion of prices in CPIH and CPI; however, there are many mitigating factors in place that reduce the level of risk. Therefore we feel that a Medium level of risk is appropriate.

Profile: High

Given the extremely wide coverage of the local price collection, there are likely to be prices collected for items which are of varying user interest. Moreover, given the very high weight of TNS data in CPIH and CPI, which are economically and politically important, it would be inappropriate to consider anything other than a high public interest profile.

Assessment: Comprehensive (A3)

Status: Achieved

Data collection is managed by TNS; however, Prices requirements are tightly specified under a comprehensive contract, which is periodically retendered. In the event of the contract being awarded to a new supplier, a dual collection would be necessary for one year to understand the impact on the quality and consistency of the data being provided. Prices Division are responsible for drawing up the sample frame and specifying the sampling methodology, whereas TNS manage the data collection. TNS’s performance against pre-specified key performance indicators is evaluated by a dedicated team within Prices Division. This is discussed with TNS at monthly operations meetings.

Quality assurance and validation procedures are applied by both TNS and Prices staff. These routines are fit for purpose, transparent and well understood.

Considering the evidence summarised above, and provided in detail in Annex A, we believe that TNS data meet the comprehensive level of quality assurance required for the production of CPIH. More detail on price collection arrangements, and quality assurance and validation procedures is provided in Annex A.

Valuation Office Agency (VOA)

Data usage

Valuation Office Agency (VOA) rental prices cover England and are used to construct indices for both private rents in CPIH and CPI, and owner occupiers' housing costs (OOH) in CPIH. In particular, the OOH index is a very large component of CPIH, and data for England account for approximately 15.14% of the weight in the headline index. The private rental index accounts for 3.25% of the weight in the headline index, the majority of which will be due to England data.

Risk: High

VOA data have a relatively high weight in CPIH, compared to other items in the basket of goods and services. Moreover, there are a number of factors that increase the risk in the use of this data.

We do not have direct access to the microdata, as the data are subject to the Commissioners of Revenue and Customs Act 2005. VOA instead supplies aggregate indices for use directly in the private rental and OOH measures. This limits the quality assurance that Prices Division can conduct on the source data.

However, the price collection is carried out by VOA rent officers who collect a purposive 10% sample of rental prices. The data source is less complex than many, and is therefore a mitigating risk factor. Moreover, there are a number of other sources of rental price data that could be considered, should VOA become unable to provide the required data.

Nonetheless, the lack of direct data access, and the high weight of VOA data in CPIH, suggests that a high-risk profile is appropriate for VOA data.

Profile: High

The OOH component of CPIH was the focus of required methodological improvements that led to the de-designation of CPIH as a National Statistic in 2014. Moreover, owner occupiers' housing costs is an area of economic importance, and the way in which these costs are measured is widely debated. For these reasons, we consider VOA data to have a high public interest profile.

Assessment: Enhanced (A2)

Status: Not achieved

Whilst we have comprehensive detail on VOA's data collection, and quality assurance and validation procedures, we do not currently have access to the microdata. Instead, VOA supply us with aggregate stratum indices. This places limitations on the level of quality assurance that Prices staff can carry out, and reduces clarity and transparency. To mitigate this risk, Prices staff have worked with VOA to implement our new methodology, set up systems, and put appropriate quality assurance and validation checks in place.

Moreover, we have put a Service Level Agreement in place to ensure continued delivery of the data, and we hold monthly supplier meetings with VOA to discuss any current or potential issues with the data supply. We also commission an external audit of VOA processes on an annual basis.

Finally, we have drafted a business case to acquire VOA microdata once the Digital Economy Bill has been passed later this year. Should the business case be accepted, we will have direct access to the microdata. This will allow us to have control over the compilation process, and fully quality assure any unusual movements in the data. This would raise the assessment of VOA data to comprehensive, as required.

More detail on the operational context, communications, and Prices and VOA data checks are provided in Annex A.

Remedial actions:

1. We have drafted a business case to access the VOA microdata once the Digital Economy Bill has been passed. We will submit this at the earliest opportunity.

5.3 Assurance level: Enhanced

We have assessed a further four of our data sources as requiring an enhanced level of assurance.

This means that we require a relatively complete understanding of the operational context in which data are collected, with an overview of sources of bias, error and mis-measurement. We also require an effective mode of communication with these suppliers and agreement for ongoing data supply. We require a relatively complete understanding of the supplier's quality assurance principles and checks. Our own quality assurance and validation checks should be proportionate and transparent, and we will communicate any risks that arise from the data.

More detail is provided for each of these four suppliers below.

Living Costs and Food Survey (LCF)

Data usage

LCF data are used to produce item level weights in CPIH and CPI. COICOP5 is the level of aggregation above item, and so LCF expenditure totals are rescaled to match the HHFCE expenditure totals at the COICOP5 level. LCF account for most of the weight at item level in CPIH and CPI (for example, the OOH item weight is taken from HHFCE data). LCF data are also one of the tools used in the annual basket update, to determine new items for inclusion and old items for removal. The data are delivered to Prices Division on an annual basis.

Risk: Low

LCF data represent all of the items in the basket of goods and services at the item level, but are not used for higher level aggregation. The data source is a survey and, although the survey design itself is complex, no other administrative data sources are used in its construction. We therefore consider it as a non-complex source. The data are collected by ONS field staff, and the survey is managed within our Social Surveys Division. If LCF data were unavailable we could consider using national accounts data instead; however, this risk is unlikely to occur.

Sample design, survey methodology and quality assurance procedures are well documented in LCF publications and, as the data are from a survey, we also have standard errors which help us to understand the accuracy of the data. One drawback of the data is that falling response rates reduce the LCF sample size and representativeness.

LCF supply the data in spreadsheet form, which can be automatically read into Prices spreadsheets. The data supply process is well established, and annual meetings are held with the LCF team.

Therefore, we consider LCF data to be low risk in the production of consumer price inflation measures.

Profile: High

Given the extremely wide coverage of LCF data, we have expenditure weights for items of varying interest amongst users. Moreover, CPIH and CPI are economically and politically important, and LCF data are used for nearly all items. The annual basket updates also tend to receive wide media interest, although LCF data are not the primary source of information for this. Therefore, it would be inappropriate to consider anything other than a high public interest profile.

Assessment: Comprehensive (A3)

Status: Achieved

The LCF team have provided detailed information on their data collection, processing, and quality assurance and validation procedures. The survey is managed by the LCF team and no other data sources are used; therefore, the information provided gives a comprehensive understanding of LCF data. Moreover, good communication mechanisms are in place with LCF, with supplier meetings held on a twice yearly basis (a planning meeting before delivery, and a review meeting after). Deliveries for CPI and CPIH are based on finalized data. There is a risk that falling response rates will introduce bias into the results; however, LCF have adopted a number of strategies to counteract this.

The LCF recently underwent a National Statistics Quality Review (NSQR), the recommendations of which are currently being delivered. The Prices delivery system was reviewed and rewritten, which has reduced the risk of manual errors.

Considering the evidence detailed in Annex A we believe that our level of quality assurance for LCF exceeds the standard required for the production of CPIH and CPI.

Mintel

Data usage

Prices Division purchases market research data from Mintel, for use in the production of some weights at and below the item level, for quality assuring unusual movements, and also for establishing new items for inclusion in the annual basket update, as well as new shops. It is hard to precisely specify the weight that Mintel data have in CPIH and CPI.

Risk: Medium

Mintel data do have quite wide coverage in the basket; however, they are used below the item level as strata weights and at item level to refine LCF weights. As with LCF data, they are subsequently constrained to COICOP5 totals. The data available from the Mintel website are drawn from a variety of sources, usually from surveys run by Mintel themselves. Their methodology, processes and quality assurance procedures are consistent and well documented. Data are generally copied into Prices spreadsheets from source.

The data are purchased on contract and, as part of this contract, Prices are allocated a designated contact. If we could not access Mintel, then it would be a straightforward matter to retender the contract and source similar data from an alternative market research company.

We assess Mintel data as being a medium risk of quality concerns. This reflects the variety of surveys used, and their relatively wide coverage in the basket.

Profile: Medium

Mintel data do not represent as broad a cross-section of the basket as HHFCE or TNS data do. This is, in part, due to the lower levels at which the data are employed, and partly the coverage. As with LCF data, they are used in the annual basket updates, and the coverage is wide enough that some items are likely to gain a wider media or user interest. For that reason we feel that a public interest profile of medium is appropriate for Mintel data.

Assessment: Comprehensive (A3)

Status: Achieved

Mintel are a well established and reputable market research company, who provide a variety of different reports drawn from various surveys and contracted agencies. Mintel have provided detailed information on questionnaire design, sampling procedure, quality assurance and validation checks, and audits. The detail provided is substantial, and Mintel's procedures are comprehensive. We are therefore satisfied that the level of quality assurance for Mintel data is appropriate for the purposes for which the data are required.

Mintel data are provided to Prices under a contract, which is renewed every 2 years. Prices have a dedicated contact who will respond to queries and concerns.

Detailed information on the operational context, communications, and Prices and Mintel data checks are provided in Annex A.

Scottish government

Welsh government

Data usage

Welsh government data are used to produce the rental price series for Wales in the private rental price index, and the OOH component of CPIH. Welsh government data are also used to produce strata weights for the Wales stratum of the OOH component in CPIH. This stratum has a weight of 0.52% in CPIH. Welsh government data are likely to represent a small proportion of the 3.25% weight for the private rents index.

Scottish government data are used in the same way, and the Scotland OOH stratum has a weight of 1.22% in CPIH. Scottish government data are also likely to represent a small proportion of the 3.25% weight for the private rents index.

Risk: Low

Both Welsh government and Scottish government data are used for very low weight components of CPIH (that is, less than 1.5%).

For rental prices, data are collected in a similar manner to those collected by VOA in England. This is a relatively simple price collection. Unlike VOA data, however, Welsh and Scottish governments provide us with the microdata, which means that we can fully quality assure the data, and maintain control over methodology and processes. As with the VOA, if these data were unavailable, we could source the information from other data providers. For both, we also have SLAs in place to ensure the delivery schedule is maintained, and we hold annual meetings with them.

For strata weights, data are typically compiled from a variety of different administrative and survey data sources, in a similar manner to the DCLG component. The dwelling counts tend to be relatively stable over time, and it would be clear if there were errors in the data. If these data sources were unavailable it is unclear what other source could be used; however, this risk is unlikely to materialise. There are dedicated contacts in place for the data delivery. There is some material published online relating to methodology, and quality assurance and validation procedures.

There is little risk of quality concerns with the rental price data, and whilst there may be slightly higher risks associated with the dwelling stock data, taking into account the very low weights ascribed to these sources, a risk profile of low seems the most appropriate.

Profile: Medium

Whilst the OOH component and private rental price indices are thought to be of wider media and user interest, there is less focus on the devolved regions. There has been much debate and focus on the OOH component; however, the overall weight of the Wales and Scotland strata in the OOH index is approximately 3% for Wales and 7% for Scotland, which is relatively small. Clearly there is some extensive interest in these indices; however, in terms of economic importance, we do not feel that anything higher than a medium public interest profile is warranted.

Assessment: Comprehensive (A3)

Status: Achieved

For private rental price data, we have comprehensive detail on the data collection, processing, and quality assurance and validation procedures for both Welsh and Scottish governments. Data collection is run by the suppliers and no auxiliary data are used. Unlike VOA data, Welsh and Scottish governments provide us directly with the microdata, which allows us to have control over the processing of elementary aggregates, and directly interrogate unusual data. Considering the evidence provided in Annex A, we believe that the level of quality assurance exceeds that required for the purposes of CPIH.

SLAs are in place with both suppliers, and annual meetings are held to discuss any data related issues.

More detail on the operational context, communications, and data checks are provided in Annex A.

Assessment: Enhanced

Status: Achieved

For dwelling stock counts, Welsh and Scottish data are pulled together from a variety of sources in a similar manner to DCLG dwelling stock counts. Both suppliers hold some form of communication with their suppliers, and use detailed validation routines to assure the data as they are compiled. All data are quality assured with reference to the time series to identify any unusual movements. We are satisfied that the procedures they have described are fit for the purposes for which they are used in CPIH and CPI, and we have a dedicated contact for both suppliers with whom we can query any unusual data points. More detail on this is provided in Annex A.

5.4 Assurance level: Basic

We have assessed the remaining data sources as requiring a basic level of assurance. This means that we require an overview of the operational context in which data are collected, and any actions taken to minimise risks. We also need to provide the supplier with a clear understanding of our requirements, and have contacts in place to report queries to. We require an overview of the suppliers' quality assurance principles and checks, and should have our own quality assurance checks in place on the data.

More detail is provided for each of these suppliers below.

Department for Business, Energy and Industrial Strategy (BEIS)

Data usage

BEIS data are used to construct weights for a number of energy items in the consumer prices basket of goods and services. In total the motor fuels items contribute a total weight of 2.58% to headline CPIH through:

- Prices for petrol (1.64%)
- Prices for diesel (0.94%)

Risk: Low

Data for motor fuels (petrol and diesel) are collected through a survey, administered by BEIS staff. The weight for motor fuels in CPIH is small (but not negligible) at 2.58%. If the data were unavailable to us, we would investigate alternative sources and, if no such sources exist, we would have to equally weight stratum level indices.

We have a dedicated contact to respond to data queries. Figures are provided by BEIS in spreadsheet form and transferred into Prices spreadsheets. Some methodology and quality assurance information is provided.

Given the low weight of BEIS data in headline CPIH and the relative simplicity of the source data, we consider BEIS data to have a low risk of quality concerns.

Profile: Low

Whilst there may be some media interest in price changes for motor fuels, this tends to be limited as regards consumer price inflation. The contribution of BEIS data to headline CPIH is not large enough to consider the economic importance of headline inflation here.

Assessment: Basic (A1) to Enhanced (A2)

Status: Achieved

BEIS data are derived from a survey conducted within the department. They have provided us with detailed information on the data collection, methodology and quality assurance procedures, which we consider to be fit for the purpose for which they are used within CPIH and CPI. These are provided in more detail in Annex A. We also have a dedicated contact for any data-related queries.

Department for Transport (DfT)

Data usage

Department for Transport (DfT) data are used in the calculation of a number of expenditure weights. In total these weights make up 5.21% of CPIH. Specifically they are used for:

- below item strata weights for used cars (1.40%), in conjunction with Glasses data
- below item strata weights for new cars (2.10%), in conjunction with Glasses data
- below item strata weights for vehicle excise duty (0.55%)
- below item strata weights for motorcycles (0.07%)
- item weights for London transport (0.25%) are constrained to COICOP5 totals
- item weights for underground fares (0.03%) are constrained to COICOP5 totals
- item weights for Euro Tunnel fares (0.04%) are constrained to COICOP5 totals
- item weights for rail fares (0.77%) are constrained to COICOP5 totals

Risk: Low

Together, DfT data constitute a small to medium weight in headline CPIH and CPI. However, there are a number of mitigating factors to consider:

- of this 5.21%, only 1.09 percentage points are used directly for item weights
- of the remaining 4.12 percentage points, 3.50 percentage points are used in conjunction with Glasses data to construct below item level strata weights
- the remaining 0.62 percentage points are used to calculate below item weights without reference to other data sources
- whilst all the data are sourced from DfT, each comes from a different DfT survey or output, we therefore seek quality assurance information for each of the components separately; however, for the sake of brevity, we consider setting assurance levels at the supplier level, taken separately, each of the components make a small to very small contribution to the total weight in CPIH and CPI

In each case, if the data were not available, we would seek alternative data sources and, in the absence of a suitable alternative, equally weight each item within COICOP5 (or item) totals. Much of the data are sourced through email contact with DfT, and either copied into Prices spreadsheet systems, or read in directly. Where item weights are being constructed, data are copied directly from tables in the latest release, and used to create a weight distribution constrained to COICOP5 totals.

Considering the various factors described above, and in particular that we are seeking a separate assurance for each series, we feel that the risk of quality concerns is low.

Profile: Low

All of the series above are of limited media and user interest. (Whilst rail fare increases are often covered by the media, this tends to be at the point when increases are announced; there is limited interest in the item index). Taken together, the series make a small to medium contribution to headline CPIH and CPI. We therefore suggest that a low public interest profile is appropriate.

Assessment: In progress

Status: In progress

Some detail on the data collection, methodology and quality assurance procedures for DfT data is available online, and they have provided us with comprehensive detail on their quality assurance, data collection, and general process for some items (Eurostar fares, rail fares, and London Transport). We are satisfied that our communications with DfT and the information provided give us a basic to enhanced level of assurance for these items. We are currently working with DfT to get the additional information required to allow us to complete this assessment.

Remedial actions:

1. We will seek further information from DfT on data collection, methodology, and quality assurance procedures to allow us to make an assessment of these data sources.

Glasses

Data usage

Glasses provide valuation data for used cars around the country. They provide these valuations for various customers (notably, car dealers, who can set their price strategy appropriately). They are a well established and reliable producer of car valuation data. The data contribute to 4.25 percentage points of headline CPIH through the following item indices:

- Glasses data are combined with Department for Transport (DfT) data to produce below item strata weights for used cars (1.40%)
- Glasses data are combined with Department for Transport (DfT) data to produce below item strata weights for new cars (2.10%)
- price data for motorbikes (0.07%)
- price data for caravans (0.68%)

Risk: Low to medium

Taken together, the contribution of Glasses data to headline CPIH is not insignificant, but it is also not large. Of this, only 0.75 percentage points is used directly at the item level, the remaining 3.50 percentage points are used below the item level in conjunction with DfT data to produce strata weights. The data source is compiled from several different sources, and so is reasonably complex. If Glasses data were unavailable, we would switch to other sources, such as used car websites, or directly from company websites.

Data are purchased via annual subscription, and queries are dealt with through regular email contact. Price data are extracted manually from the website, whereas expenditure data are received in spreadsheet form, which can be read directly into Prices spreadsheet systems. There is also a great deal of information on their methodology and processes available online; however, detail of their quality assurance procedures is not provided.

Considering the small to medium weight, how the data are used, and the existing arrangements, we feel that Glasses data merit a low to medium risk profile.

Profile: Low

Indices for used and new cars, and for motorbikes and caravans are of little user and media interest, and their overall contribution to CPIH is not large enough to consider their contribution to the headline index relevant. Therefore we make an assessment of low public interest profile for Glasses data.

Assessment: Basic (A1)

Status: Incomplete

Glasses data are compiled from a variety of sources. The data are purchased through a yearly subscription, and a help desk number is provided for queries. There are some concerns over communication, as Glasses have not yet shared their quality assurance and validation procedures with us. However, there is a great deal of information available publically through their website. There was also a lack of communication from the supplier when data transfer moved from CD to online. Checks carried out by members of staff within Prices Division are comprehensive, and queries are raised through the help desk. Further detail is provided in Annex A.

Remedial actions:

1. Establish better lines of communication with Glasses, by seeking a dedicated point of contact within the company
2. Continue to request information on Glasses' quality assurance procedures

International Passenger Survey (IPS)

Data usage

IPS data are used to construct strata weights below the item level for foreign holidays. They are used in conjunction with Mintel data. Foreign holidays make up 2.55% of the weight in headline CPIH.

Risk: Low

The data have a low, but not insignificant weight in headline CPIH. IPS data are collected through a survey, supplemented with some administrative data. Nonetheless, the data structure is relatively straightforward compared to some other sources. Moreover, as the basis of IPS data is a survey, their properties are better understood than data which are compiled from many administrative sources. The data are collected, processed and compiled by our staff within Social Surveys Division. If we did not have IPS data, we would instead use our market research data and, failing that, below item level indices would be given equal weight. The methodology, and quality assurance and validation procedures are well documented.

Profile: Low

Foreign holidays are of limited media and user interest. They also have a relatively low weight in CPIH, which is of greater economic importance. Therefore we will consider IPS data to have a low public interest profile.

Assessment: Basic (A1) to Enhanced (A2)

Status: Achieved

IPS data are largely produced via survey, which is run by the IPS team; however, some auxiliary administrative sources are also used. IPS have provided detailed information on the quality assurance procedures applied to their source data and their outputs, as well as methodology and processing. We are satisfied that the procedures described are fit for the purposes for which they are used in CPIH and CPI. Further details are provided in Annex A.

Consumer Intelligence

Higher Education Statistics Authority (HESA)

Home and Communities Agency (HCA)

Inter-Departmental Business Register (IDBR)

Kantar

Moneyfacts

Data usage

Consumer Intelligence data are used to get prices for house contents insurance and car insurance. The combined weight for these items is 0.43%.

HESA data are used to calculate strata weights (below the item level) for University tuition fees for UK and international students. The combined weight for this item is 1.05%.

HCA data are the source of rental price data for registered social landlords. The weight for this item is 1.34%.

IDBR data are used to derive below item strata weights for boats. The weight for this item is 0.29%.

Kantar data are used to calculate below item strata weights for a number of digital media items: internet bought video games, DVDs, Blu-Rays and CDs, and downloaded video games, music and e-books. The combined weight for these items is 0.36%.

Finally, Moneyfacts data are used as the source of price information for mortgage fees. The weight for this item is 0.12%.

Risk: Low

All of the data sources listed above feed into items with a very low weight in CPIH, generally less than 1.5%. As such their impact on headline CPIH or CPI will be minimal. Should any of these sources of data become unavailable, there is a clear contingency for each:

- Consumer Intelligence: Create a smaller sample, based on price quotes from comparison websites
- HESA: Equally weight courses and institutions below the item level
- HCA: Investigate the use of alternative sources of price data
- Kantar: If finances are not available to purchase the data, Mintel data can be used instead
- Moneyfacts: Collect prices from individual company's websites

Kantar data are collected through the use of a survey, and Consumer Intelligence data are scraped from supplier websites. IDBR data are more complex, being compiled from 5 different data sources, and HESA data are compiled from all Higher Education institutes across the UK. We are not aware of the sources for Moneyfacts data. All of the data are manually fed into spreadsheets, which use formulae to derive the subsequent price index.

A contract is in place to receive Consumer Intelligence data, and regular supplier meetings are in place. MoneyFact data are acquired through an annual magazine subscription, and Kantar data are purchased annually on an ad-hoc basis. Kantar also provide a dedicated contact. There is no contractual agreement in place for either HESA or HCA; data are instead sourced through direct email contact with the supplier. The IDBR team is based within ONS, so no contract is in place. None of these arrangements are out of keeping with the weight accorded to these items in the basket.

There are some risks associated with these data; however, given their negligible impact on headline CPIH or CPI, we do not feel that the risks associated with use of these data sources merit anything higher than a low level of risk.

Profile: Low

The above data sources are used in the construction of very specific low-level item indices. They may be used to capture the price element of the index, or they may be used for below item-level strata weights. They will generally be combined with prices or strata weights to create the particular index.

With the possible exception of tuition fees, none of the item indices are considered to be of wider user or media interest, and are certainly not politically or economically sensitive. They are generally of niche interest and are politically neutral. Tuition fees can be of interest following a major change; however, such changes are rare and HESA data are only used below the item level. As described under risk, their contribution to CPIH and CPI, which are considered to be economically important and market sensitive, is very small (less than 1.5%) and, as such, their impact on the headline figures is negligible.

Assessment: Basic (A1)

Status: Achieved

Consumer Intelligence is a well established and reputable market research company, who send us a sample of insurance quotes. We have a dedicated contact; however, at present we have been unable to obtain further quality assurance information as the contact has not responded.

HESA data are sent to Prices Division in an Excel spreadsheet. There is a data sharing agreement in place to access the data, and a dedicated contact. Quality assurance procedures are well documented by HESA, and all input data sources are listed.

HCA rental prices for registered social landlords are obtained through direct email contact with the supplier. We have engaged with HCA, who have provided an overview of their data collection process, and quality assurance and validation procedures, which we consider to be fit for the purpose for which they are used in CPIH and CPI.

IDBR have provided us with information on their data collection, methodology and quality assurance procedures. Data are compiled from a number of sources and IDBR's procedures for validating these sources are clear.

Kantar is a well-established and reputable market research company. Data collection is administered through a longitudinal survey, and the survey methodology and quality assurance procedures have been communicated to us. We consider these to be fit for the purpose for which they are used in CPIH and CPI.

Moneyfacts are a price comparison company, who collect data from websites. We collect the data through a monthly magazine subscription. There is no dedicated contact, so contact details must be sought from the Moneyfacts website. We have some information on the coverage and data collection; however, quality assurance information is not readily available for Moneyfacts. Prices collection and quality assurance procedures are robust. For example, we have often checked extreme movements against company websites, and found the data to be correct.

More detailed information on all of the above is available in Annex A.

Remedial actions:

1. Clarify contact details for Consumer Intelligence
2. Seek further quality assurance information from Consumer Intelligence
3. A dedicated contact for Moneyfact should be established and kept current
4. Further detail of Moneyfacts' quality assurance procedures should be sought

Websites

Direct contact

Brochures, reports and bulletins

Data usage

Price collection from websites is used to collect prices for many of the items which are not sourced through the local price collection (currently conducted by TNS). Website collections account for approximately 5% to 10% of the weight in CPIH.

Price collection through direct contact (typically by phone or email) accounts for approximately 5% to 10% of the weight in CPIH, and is used for items which are not collected locally or through websites.

Price collection from brochures, reports and bulletins accounts for approximately 1.5% to 5% of the weight in CPIH, and is used for items not collected through local collection, websites or direct contact.

These price collections are referred to as "central" collections.

Risk: Low

Whilst these collections have a small to medium, or medium weight in CPIH, there are a number of factors that reduce the risks substantially:

- All of the price collections are conducted in-house by staff in Prices Division. This gives us complete control over the process
- For all of these collections, there is a very clear and achievable course of action, should a data source become unavailable :
 - if a retailer's website becomes unavailable, then a new website can simply be identified, this is analogous to a shop closing in the local price collection, where we would simply find a new shop to collect the data from; it is extremely unlikely that more than one or perhaps two websites would close down in a given month, and so this is unlikely to cause issues for price collection
 - if we are unable to continue collecting from a direct contact supplier then, again, we can simply identify a replacement supplier to collect the prices from
 - should we be unable to source appropriate brochures, reports or bulletins, then we could simply identify alternative internet-based sources instead; many of the sources are purchase on annual subscription, so this provides some additional security for ongoing collections

The nature of these collections means that Price quotes will need to be manually entered into Prices processing systems. Robust quality assurance and validation procedures are in place for these processes, and are described in more detail in Annex A.

Profile: Low

None of the centrally collected items are of wider media or user interest, and are not economically or politically important. Whilst taken together their contribution to headline CPIH is large, they actually represent specific collections for many different items. Therefore we assign a low public interest profile to centrally collected data.

Assessment: Basic

Status: Achieved

The assessment of these sources is focussed on Prices own procedures, as these sources are essentially an in-house data collection conducted by Prices staff. This means that we are effectively both the supplier and the producer. We have robust quality assurance checks in place, and our data collection process is recognised under ISO9001: 2015, and supported by in-house staff training. Further information on these is presented in Annex A.

6 . Action plan

In the previous sections we have considered quality assurance for all data sources in our consumer price inflation measures. We assessed the required assurance levels by considering the risk of quality concerns for each data source, and the public interest profile of the item they are used to calculate. We then conducted the assessment based on four practice areas: operational context and data collection, communication with data supply partners, quality assurance (QA) checks by the supplier, and our own QA investigations. This information is detailed in Annex A.

Of the data sources we investigated, there are several that need further work to reach the level of assurance we are seeking.

For Valuation Office Agency (VOA) data, we do not have access to the microdata. Users should be aware that we do not have complete control over the production process, and this limits our ability to quality assure the data. However, we have put several mitigation strategies in place, such as working with VOA to develop the methodology, processing and quality assurance, and developing a business case to access the data once the Digital Economy Bill has been passed. Moreover, the data source is very strong, with a sample of approximately 500,000 price quotes annually, and our research has shown that it is broadly [comparable with other sources](#). We are therefore satisfied that users can be confident in the VOA data used to construct the owner occupiers' housing costs (OOH) component of Consumer Prices Index including owner occupiers' housing costs (CPIH).

For Household Final Consumption Expenditure (HHFCE), we would like a fuller understanding of how quality assurance has been applied to the source data used to construct expenditure estimates. HHFCE estimates are based on a complex array of data sources, and users should be aware that these are not necessarily fully understood. HHFCE data, however, remain the most suitable source of weighting information for consumer price indices, following international best practice. Their quality assurance and validation procedures should be comprehensive enough to identify any issues in the source data, and we have a good understanding of the data, given that they are also produced within ONS.

Similarly, for the Department for Communities and Local Government (DCLG), we would like a fuller understanding of the quality assurance that has been applied to the source data and how that impacts on dwelling stock estimates. Dwelling stock counts are also based on a number of different data sources and, again, users should be aware that these are not necessarily fully understood. The data themselves, however, are used below the item level in conjunction with VOA data to estimate strata weights. They tend to be very stable between years, and so issues in the data are easily identifiable. DCLG's quality assurance and validation procedures are comprehensive and should identify any data issues, as are our own.

Finally, there are a number of data sources for which we have sought a basic level of assurance, and for which additional quality assurance information has been requested but, as yet, has not been provided. Moreover, contacts for some of these sources are out of date or unknown. We will continue to work with suppliers to better understand their processes. Users should be aware that our understanding of the data is incomplete; however, the risk to headline CPIH or CPI is minimal, as reflected in the basic assurance requirement.

To address these shortcomings, we will carry out further steps to improve our quality assurance. All outstanding actions are summarised below (Table 5), with details on what actions we intend to take to rectify them.

Table 5: Action plan

Provider	ONS Action
DLCG	Seek to understand DCLG's departmental approach to QAAD development, and where dwelling stock count data fits into this process Seek to establish a more robust delivery schedule, including annual supplier liaison meetings, and a Service Level Agreement for delivery
HHFCE	HHFCE are in the process of completing a QAAD assessment of their data sources. They aim to complete their QAAD assessment by autumn 2017.
VOA	We have drafted a business case to access the VOA microdata once the Digital Economy Bill has been passed. We will submit this at the earliest opportunity.
DfT	We will seek further information from DfT on data collection, methodology, and quality assurance procedures to allow us to make an assessment of these data sources.
Glasses	Establish better lines of communication with Glasses, by seeking a dedicated point of contact within the company Continue to request information on Glasses' quality assurance procedures
Consumer Intelligence	Clarify contact details for Consumer Intelligence Seek further quality assurance information from Consumer Intelligence
Moneyfacts	A dedicated contact for Moneyfact should be established and kept current Further detail of Moneyfacts' quality assurance procedures should be sought
Various	Set up firmer ongoing communication mechanisms
Various	Establish firmer data delivery agreements

Source: Office for National Statistics

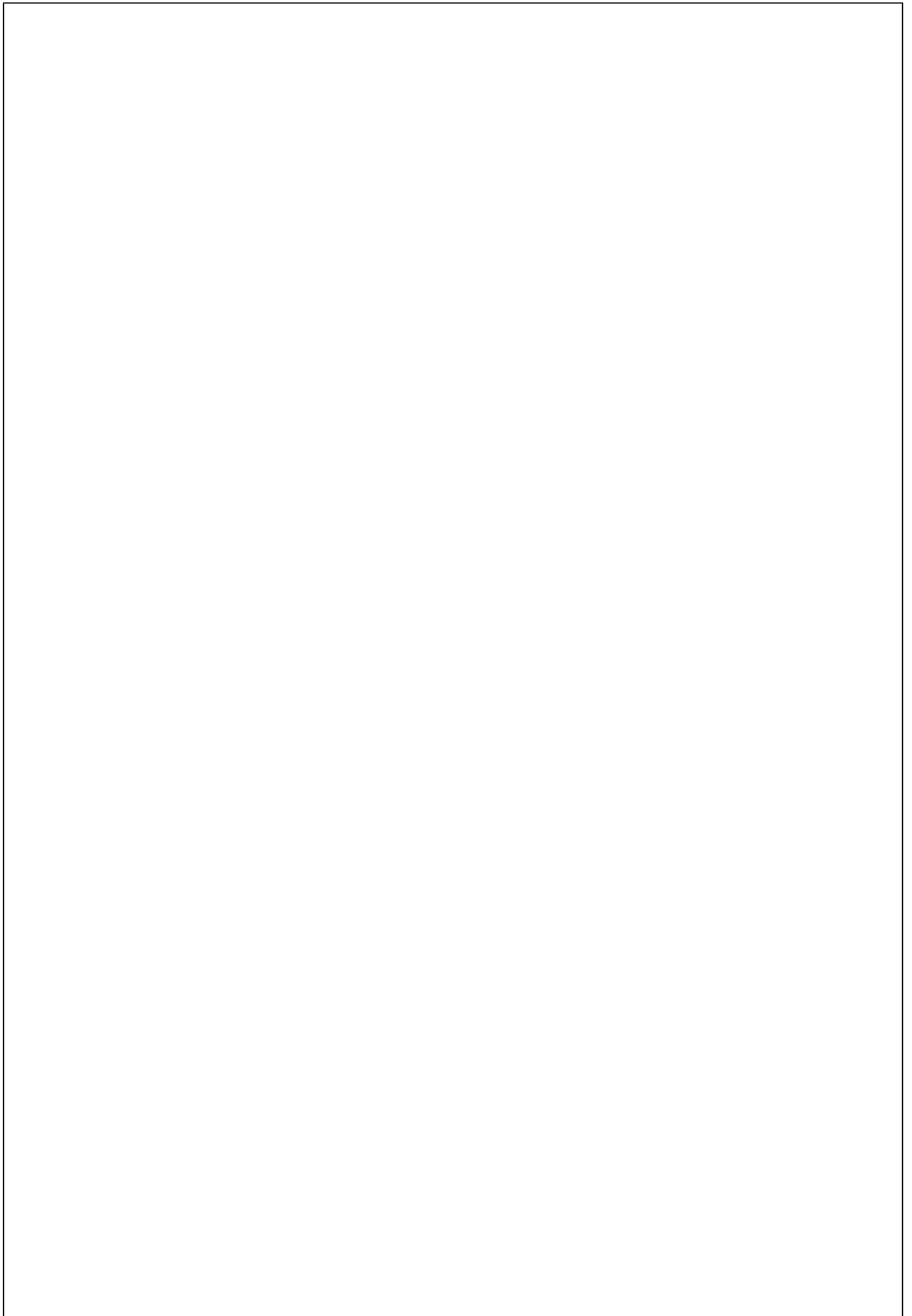
This version of the consumer price statistics QAAD is intended to act as a progress update. Over the next few months we intend to continue engaging with our data suppliers and, where appropriate, put in place firmer ongoing communications mechanisms and data delivery agreements. We will aim to publish an update to this QAAD in summer 2017. Importantly, this QAAD is not intended to serve as a final record of quality assurance. We view supplier engagement and feedback as an ongoing process, which we will continue to follow. We therefore intend to publish a review to this QAAD every 2 years.

7. Annex A: Assessment of data sources

1. Department for Communities and Local Government (DCLG)

Practice Area 1: Operational Context and Data Collection

Department for Communities and Local Government (DCLG) dwelling stock estimates are used to calculate strata weights, used to mix adjust the housing component of Consumer Prices Index including owner occupiers' housing costs (CPIH) to reflect the owner occupiers' housing costs (OOH) market. This is taken from the statistical bulletin "Dwelling stock estimates" which is updated annually. DCLG do not collect data directly for the production of the release, instead drawing on a range of data sources to compile a set of statistics on the total number of dwellings and the tenure profile of the stock. Each of these data sources is set out in Table x below, along with strengths and weaknesses of the data and DCLG's own assessment of the source.



Source or Type	Use	Strengths	Weaknesses	Continuing improvement	Why “fit for purpose”
<p>Population Censuses 2011 and 2001: dwelling counts</p> <p>Census</p>	<p>Baseline for total stock by local authority</p>	<p>1. Consistent, comprehensive coverage from national to sub local authority level</p> <p>2. The 10-year estimates fit well with stock projected by annual net additional dwellings</p>	<p>1. Infrequent</p> <p>2. Known undercount in 2001</p>	<p>Working with ONS on the potential use of Annual Population Survey and Administrative Data</p>	<p>1. ONS approved an adjustment for the 2001 error</p> <p>2. Sense checked with other sources and found consistent</p> <p>3. Consistent estimates</p> <p>4. Alternatives have inconsistencies</p> <p>Risk</p> <p>Error – minimal</p> <p>Impact – high</p>
<p>DCLG net additional dwellings</p> <p>Annual statistical return from local authorities</p>	<p>To increment local authority total stock estimates by net gains and losses to stock on an annual basis</p>	<p>1. Provided by local authorities</p> <p>2. Digital return, validated by DCLG</p>	<p>1. Does not allow revisions to previous data</p>	<p>Considering allowing revisions for earlier years</p>	<p>1. Public scrutiny, as the main measure of housing supply</p> <p>2. National Statistic</p> <p>3. Comprehensive quality assurance and sense checks by DCLG</p> <p>Risk</p> <p>Error – low</p> <p>Impact – medium</p>

<p>Local authority housing statistics</p> <p>Annual statistical return from local authorities</p>	<p>Local-authority-owned dwellings for rent</p>	<p>1. Statistical return provided from local authorities that own and manage the stock</p> <p>2. Digital return validated by DCLG</p> <p>3. High response rate (100% in 2015 to 2016).</p>		<p>Continuing communication with local authorities</p>	<p>1. Best measure of local authority stock</p> <p>2. National Statistic</p> <p>3. Comprehensive quality assurance and sense checks by DCLG</p> <p>Risk</p> <p>Error – low</p> <p>Impact – low</p>
<p>Homes and Communities Agency (HCA)</p> <p>Statistical return from Private Registered Providers (PRPs)</p>	<p>Local-authority-level estimates of PRPs' stock</p>	<p>1. PRP statistical returns on the stock they manage</p> <p>2. Digital return validated by DCLG</p> <p>3. Complete response from larger PRPs (1000+ dwellings)</p> <p>4. HCA is an executive non-departmental public body sponsored by DCLG</p>	<p>94% response from smaller PRPs, corrected by weighting</p>	<p>Regular communication with HCA about their stats</p>	<p>1. Comprehensive quality assurance by HCA</p> <p>2. National Statistic</p> <p>3. Weighting for non-response as advised by ONS and DCLG</p> <p>Risk</p> <p>Error – low</p> <p>Impact – medium</p>

<p>Labour Force Survey (LFS)</p> <p>Sample survey</p>	<p>To provide regional level split of private dwellings into owner-occupied</p>	<p>1. Established ONS survey</p> <p>2. LFS is triple quality assured by ONS, the UK data archive and DCLG</p>	<p>1. Covers occupied dwellings only</p> <p>2. Year-on-year estimates of private rented accommodation can be variable</p> <p>3. Standard errors not produced for private rented or owner-occupied split</p>	<p>1. Informed by LFS quality documentation</p> <p>2. Working with ONS on comparison with Annual Population Survey</p>	<p>1. Large sample survey (40,000 households per year)</p> <p>2. Comprehensive quality assurance and sense checks by DCLG</p> <p>3. Variable data are smoothed</p> <p>Risk</p> <p>Error – low</p> <p>Impact – medium</p>
<p>English Housing Survey (EHS)</p> <p>Sample Survey</p>	<p>Vacancy estimate for private rented sector</p>	<p>1. Established survey</p> <p>2. Standard errors documented</p> <p>3. Well regarded survey running continuously for 50 years</p>		<p>Continuing dialogue with EHS</p>	<p>1. Large sample size (over 13,000 per year)</p> <p>2. Comprehensive quality assurance and sense checks by DCLG</p> <p>3. Best source of vacancy in the private rented sector</p> <p>Risk</p> <p>Error – low</p> <p>Impact – medium</p>

The production process is conducted in Excel and SPSS, with data being transferred to spreadsheets and formula used to calculate the dwelling stock estimates.

The total for the private sector is first calculated by deducing the full counts for the social housing sector, using data from local authorities (LAs) and private registered providers (PRPs).

Data are collected from LAs via an annual form-based data return, and PRPs provide data annually through a web-based data capture system called NROSH+.

To ensure the consistency of returns, DCLG provides [guidance notes](#) for the LA housing statistic annual return. These guidance notes are updated annually. Moreover, the online form has interactive validation that alerts the

user to invalid and implausible values. There is a further validation check after the data have been received, which will follow up with LAs as necessary. They also hold regular supplier communications, as discussed under practice area 2.

Detailed guidance is also provided for PRPs and helpdesk support is available. Similar validation checks are used to identify implausible or invalid entries and manual checks are carried out once the data are received. The robustness of validation procedures is ensured through random spot checks on 10% of the returns.

The private stock is then split into owner-occupied (OO) which we use for CPIH, and private rental sector (PRS), which we use for the Index of private housing rental prices (IPHRP). There is no direct measure due to the difficulty of collecting the information and the fluid interchange between the two parts.

An estimate of PRS is calculated using information from the Labour Force Survey (LFS) and English Housing Survey (EHS). Estimates are first taken from the LFS and smoothed using a 3-year weighted average. The PRS is then adjusted by the occupancy rate, which is calculated as one minus the EHS vacancy rate. The EHS vacancy rates are taken from the most recently available survey information.

The OO tenure is then calculated by deducting the PRS, local authority, PRP and other public sector values from the total stock.

All data for this release have been previously published in Excel/SPSS formats. Therefore, the production process is in Excel and SPSS and involves the collation (for example, transferring data from spreadsheets) and calculation of dwelling stock estimates (using formulae). To mitigate risks, this is all comprehensively quality assured in accordance with the Department Quality Assurance Procedures and Toolkit. The methodology is fully documented in the release.

Going forward, DCLG has plans for a new IT system for all admin data collection, called DELTA (incremental delivery by summer 2018) and afterwards a new analytics system called DAP, which would help mitigate these risks.

Practice Area 2: Communication with data suppliers

DCLG hold frequent communication with these data suppliers. There are quarterly Central Local Government Information Partnership (CLIP) meetings with representatives from local authorities. CLIP-planning focuses on the New Supply of Housing and CLIP-housing on the Local Authority Housing statistics. There is an annual meeting with the Greater London Authority and an annual conference with representation from all 32 London Boroughs. There are additional ad-hoc visits to local authority data suppliers and the devolved administrations. Quarterly meetings are held with other data suppliers, including the Office for National Statistics, the National House Building Council and the Homes and Communities Agency. Additionally there are frequent meetings with the English Housing Survey, user engagement events and regular written communication with data suppliers.

Users and uses

Dwelling stock estimates are used as evidence in policy making by central and local government. The data is used in the development and production of other government statistics such as the English Housing Survey and the ONS.

Since the previous publication of this QAAD we have spoken to several users of DCLG data.

DCLG dwelling stock counts form part of the Structural Housing Indicators (SHIs), which are compiled for the Working Group for General Economic Statistics (WG GES), chaired by the European Central Bank (ECB). Its membership is comprised of delegates from across the 28 EU member states.

Member states' SHI contributions are used to form an EU and euro area level aggregate, weighted according to the share of dwellings in each country. The principle purpose of the SHI data is for internal ECB use; however, ECB analysis may be included on an ad hoc basis in reports by the European Systemic Risk Board, financial stability reports and so on.

The Bank of England (BoE) use DCLG dwelling stock data to investigate shifts in the tenure share and modelling the outlook for tenure, particularly in light of the fall in owner occupation. They are also interested in looking at housing supply trends, house prices and the interactions between household and population measures. Charts from the analysis will often feed through into various BoE publications on an ad hoc basis, such as the biannual Financial Stability Reports and policy statement by the Financial Policy Committee on buy-to-let housing. The analysis in the reports will generally be cross-checked against the original data source.

Our Housing Analysis Team is currently developing local authority (LA) level estimates of dwelling stock by tenure². They benchmark Annual Population Survey (APS) LA-level survey estimates to DCLG regional tenure totals. The data are not yet published; however, the Housing Analysis Team is hoping to produce the data as experimental statistics within the next few months. The Welsh Government currently publishes LA-level tenure estimates for Wales, and so this new publication would complement the Wales statistics with a set of English statistics. The longer-term aim is for these figures to become a National Statistic.

All of these users carry out (or have carried out) plausibility checks on the data. This includes cross-referencing the data with similar measures or, in the case of ECB, against movements in similar economies. The data are also typically compared against the previous time series to establish unusual movements in the data. Users will also check that the data are meaningful.

Dwelling stock estimates are used by finance and investment industries, for example to develop a picture of demographic trends.

There are several other published statistics which attempt to measure the same concept:

[Net Supply of Housing](#)
[House building starts and completions](#)
[Affordable housing supply](#)
[New homes bonus](#)

Practice area 3: Quality assurance principles, standards and checks by data suppliers

Quality assurance and validation checks

There are automatic validation checks at entry for the Housing Flows Reconciliation (HFR) forms. In 2016, 5.46% of Local Authorities (LAs) who return an HFR failed these validation checks. This is fairly typical of other years. If an LA's HFR is flagged during validation, DCLG staff will contact the LA to confirm the accuracy of the return, and the data are amended if necessary. All of the queries in 2016 were resolved. Validation checks include:

- (x bedroom) social rent dwelling stock and (x bedroom) affordable rent dwelling stock cannot exceed total number of (x bedroom) dwellings
- Total number of new builds cannot exceed total stock
- Selling price for all dwellings cannot be less than the selling price for all flats

There is then a thorough QA check of all data, with any issues being discussed and resolved with Local Authorities before publication. Quality assurance procedures are consistent with departmental guidelines on quality assurance checking. Checks include:

- querying unlikely values with the data provider
- range and outlier checks
- time series comparisons
- plausibility checks for frequency counts and other relevant statistics
- spot checks to ensure that formulae have been used and copied correctly

The Dwelling Stock Estimates process and quality assurance checks are reviewed annually; for example, in 2013 to reflect the release of new Census 2011 data. However the process and quality assurance for the underlying data sources collected by the department are also reviewed throughout the year. More information can be found in the most recent release of the statistic:

[Net supply of housing](#)

There is then a thorough QA check of all data, with any issues being discussed and resolved with Local Authorities before publication. Quality assurance procedures are consistent with departmental guidelines on quality assurance checking. Checks include:

- data collection: Have any unlikely values been queried with the data provider
- raw data: Do frequency counts and other relevant statistics for each variable appear plausible?
- data manipulation and analysis: Have spot checks been done to ensure that formulae have been used and copied correctly?
- release preparation: Is the pre-release access list up to date?

More information can be found in the most recent release of the statistic.

Missing data or imputation

The dwelling stock estimates for England use the existing 2001 census and 2011 census as a baseline. Wales is similar and added a measure of net supply for each intervening year. Scotland use council tax data for dwelling counts.

It is estimated that the dwelling count from the 2001 census contains an undercount for England of approximately 60,000 dwellings. There is a wide margin of error around this estimate of the undercount, and our methodologists do not recommend that it should be used as a basis on which to revise the census count. For this reason, and to maintain consistency with published census figures, the dwelling stock estimates in this series continue to use the existing 2001 census and 2011 census count as a baseline.

In Scotland council tax data includes certain dwelling types which are not included in the count for the rest of the UK, and evidence suggests that this increases the Scottish dwelling stock estimates by less than one %.

Imputation is required in England for individual local authority districts, accounting for around 1% % of annual net supply. The imputation used data from all local authorities that finalised their 2015 to 2016 Housing Flows reconciliation return to calculate a ratio of the number of house building completions to the new additions figure.

This ratio was then applied to those that did not submit a HFR return. There was a 99% response rate in 2015 to 2016, so there were only four local authorities that required this imputation. This method of imputation should not lead to any positive or negative bias in the overall figures.

Review of processes

A review of the methods and data sources used to produce estimates for dwelling stocks was conducted in 2009 by the Office for National Statistics. A key finding was that the existing method remains the most suitable.

The processes and quality assurance checks for "Dwelling Stock estimates" are reviewed annually. The processes and quality assurance for the underlying data sources are reviewed through the year.

There is ongoing work with ONS to investigate extending the methodology to local authority estimates.

Revisions policy

There are two types of revision, covered by a policy which has been developed in accordance with the Code of practice for official statistics and the Local Government Revisions policy.

Non-scheduled revisions

Where a substantial error has occurred as a result of the compilation, imputation or dissemination process, the statistical release, live tables and other accompanying releases will be updated with a correction notice as soon as is practical.

Scheduled revisions

Scheduled revisions for the dwelling stock estimates are dependent on revisions to the Net supply of housing statistics. Information on the revisions policy of those statistics can be found in the most recent release of those statistics at the following link.

Additionally, the dwelling stock estimates are calibrated against the census dwelling count on its release every ten years.

Following the 2011 census, the annual figures for 2002 to 2011 were adjusted with any difference spread evenly across the 10 years. It amounted to around 16,000 extra dwellings per year at the England level. These were not evenly spread across districts.

Practice area 4: Producers' quality assurance investigations and documentation

If there are large shift in the most recent years data, then Prices division would query the reason for this with DCLG. Once the averages have been combined, Prices will look at how the weights compare to previous years. This information is provided to the CPIH production team who use it to create the item weights for England, Wales, Scotland and Northern Ireland. At this stage the item weights are also compared to previous years.

2. Household Final Consumption Expenditure (HHFCE)

Practice area 1: Operational context and admin data collection

Household Final Consumption Expenditure (HHFCE) data are used extensively in the production of Consumer Prices Index including owner occupiers' housing costs (CPIH). There are 36 different data sources which are used to construct the statistic; HHFCE have provided a list of all data suppliers and details of the information provided.

	Name	Type of source	Detail
1	Association of British Insurers	external	Insurance data annual and quarterly for all types of insurance (excluding life)
2	OFCOM	external	Communications services data- quarterly

3	BSKYB	external	Satellite subscription charges - quarterly
4	OFWAT	external	Water and sewerage services - Annual
5	Scottish Water	external	Water services - Annual
6	DECC	external	Gas electricity and motor fuels- quarterly data, first and second estimate at M2 and M3 respectively
7	DCLG	external	Housing stock –Annual
8	VOA	external	Housing rental values - Annual
9	Tourism	internal	Tourism imports and Exports – data supplied at M2 and M3
10	LCF	internal	Feeds into many HHFCE expenditure categories – Quarterly – initial estimates in time for M3 each Qtr, with re delivery of previous qtr. Annual redelivery of all 4 qtrs in line with data used in LCF publication.
11	ABS	internal	Used to benchmark RSI data -Annual at BB
12	RSI	internal	Feeds into many semi durable and durable goods items of expenditure, Quarterly at M2 and revised at M3
13	Finco's	internal	Life insurance – quarterly
14	FISIM	external	Financial Services - quarterly
15	Bank of England	external	Via FISIM
16	Population and public policy	internal	Mid-year population, births and deaths
17	CPI	internal	Price indices - Quarterly
18	GFCF	internal	Removal data - quarterly
19	CAA	external	Number of passenger air miles- quarterly
20	Transport for London	external	Underground expenditure -
21	IPS	internal	Air and sea travel expenditure-quarterly
22	HMRC	external	Data on Tobacco, Alcohol, gambling, Monthly and quarterly data
23	Department for Transport	external	Sea transport number of passengers, buses deflator. Bus fares, incl concessions
24	Gambling Commission	external	Gambling data - Annual
25	Camelot	external	Lottery data – sales and prices – quarterly and bi- annually
26	ONS - Vital statistics	internal	Midyear population, births and deaths
27	Office of rail and road	external	Rail and Road transport passenger Km prices - quarterly
28	Glass's	external	Car prices - quarterly

29	CGA	external	On trade alcohol prices and volumes - quarterly
30	A C Nielsen	external	Off trade alcohol prices and volumes- quarterly
31	Crime survey for England and Wales	external	Drug user numbers - annual

Processes

HHFCE have provided a detailed flow diagram showing the process involved in producing the statistic. We are sufficiently confident that this shows the appropriate processes and quality assurance steps. Forecasting is used for annual data deliveries for the periods used in construction of the Consumer Price Indices. Quarterly data is often informed by other, short term sources which are benchmarked to the annual deliverers. This is either forecasted or the short term source continues to be used until the data is available.

Practice area 2: Communication with data supplier partners

Internally, HHFCE hold a quarterly Prices Stakeholder board. There is a Living Cost Food (LCF) steering group, which includes informal conversations, particularly around deliveries. There is also a steering and user group for the International Passenger Survey. Additionally regular and ongoing conversations are held with Cross-National Accounts suppliers.

External suppliers from whom data specifically generated for HHFCE are contacted regularly concerning the quality and timeliness of the data. Other suppliers where data is publically available are contacted if the publication changes or if the quality of data requires confirmation.

Users and uses

HHFCE statistics are used regularly by policy departments working on both the wider economy and particular industries. The total estimate of household expenditure is an important indicator for the wider economy because household expenditure accounts for 60% %of gross domestic product (as measured by expenditure). The components of total household expenditure or Classification of Individual Consumption by Purpose (COICOPs) are useful for Government Departments interested in particular industries, for example food.

Analysts from HM Treasury use HHFCE estimates to understand the changing expenditure patterns across the economy, for example on housing. Her Majesty's Revenue and Customs use the information contained within the household expenditure estimates to analyse the tax expenditure on alcohol and tobacco products. The Department of Culture Media and Sport uses household expenditure estimates to monitor spending in their areas of responsibility: arts, broadcasting, the press, museums and galleries, libraries, sport and recreation. The Home Office uses household expenditure estimates for analysis related to crime and the economy. In March 2011 the Household Expenditure team ran a month-long consultation using Survey Monkey to better understand the needs of Consumer Trends users. The consultation was also publicised on the Royal Statistical Society website. An analysis of the survey results was published on 28 June 2011. HHFCE conforms to the European System of Accounts 2010 and the System of National Accounts 2008.

Practice area 3: Quality assurance principles, standards and checks by data suppliers

Quality assurance

HHFCE have provided a comprehensive list of quality assurance and validation checks which are completed on their outputs. These include:

- systems: Check previous round data has been archived in systems
- check there have been no failures in the local system
- check QA graphs and revisions spreadsheets have been uploaded correctly, and spot checking data within systems (one staff member to complete other to check)
- after balancing: Update briefings if necessary

- publication day: Check Time series data is consistent with publication
- more generally, adjustments can be made at a number of points in the process, depending on the source of the issue and how the change needs to be applied

Data is checked to ensure it appears in line with past deliveries; where it is not, the suppliers are contacted to confirm the variations. Analysis tables are used by COICOP input to analyse data when it has been delivered directly into HHFCE systems. A staff member completes the upload of data and checks that it looks in line with previous quarters. Further inputs are checked in the local system where possible.

Missing data or imputation

Some of the data used to produce estimates of HHFCE for Consumer Trends are only available on an annual basis. This means that some quarterly estimates are estimated by interpolation between releases of data. Generally each of the data sources have limitations, which are sometimes statistical such as missing units or under coverage, and other times are conceptual in that they do not quite measure what is required. Adjustments are made in these instances either by referencing a comparable data source or through the balancing process.

HHFCE use three key methods to compensate for missing data. Adjustment or processing at source is completed using the methodologies of the survey sources. Various data sources are combined where available to cover areas not captured by a single source. This processing or modelling is completed within HHFCE and is also used to make conceptual adjustments to more current or lower level sources. As part of the expenditure measure of gross domestic product (GDP), HHFCE is also balanced against other measures of GDP each quarter; additionally supply-use balancing is conducted as part of the annual round. These comparisons against other sources are used as a sense check for whether movements are being appropriately captured. Further adjustments such as outliers are also made as necessary.

Revisions policy

HHFCE have provided a link to their published revisions policy. This is part of [the wider revisions policy for National Accounts](#).

Practice area 4: Producers' quality assurance investigations and documentation

For a general outline of QA procedures by us, which applies to HHFCE, please refer to Annex B

3. TNS

Practice area 1: Operational context and admin data collection

The price collection for the Consumer Prices Index (CPI) and the Retail Price Index (RPI) is currently undertaken in two parts:

- central collection – by teams within Prices Division
- local price collection – under contract by Kantar TNS

The current contract for the local price collection began in February 2015 and has recently been extended to January 2020. This contract has been awarded via full competitive tender. The service is specified in detail in the contract.

Prices for approximately 520 items are currently collected in each of the locations around the UK in approximately 20,000 outlets. The total number of prices collected in each location is about 850. This is because, for some items, more than one price is collected. In the current "basket" 106,000 price quotations are allocated for local collection each month. Price data, together with additional metadata, are recorded on hand-held devices. The data are then uploaded to the Kantar TNS system and are subsequently transferred electronically to ONS. Quality checks are carried out by Kantar TNS prior to the data being transferred. A sample of the local price collection is thoroughly audited each month by ONS employees.

Practice area 2: Communication with data supplier partners

Performance review meetings are held on a monthly basis between ONS and Kantar TNS. Performance against the key performance indicators is discussed along with any issues arising from the collection that month,

developments affecting the collection and schedules for future collections. Twice a year a strategy meeting is held to discuss major changes to the collection, suggestions for improvement, variations to the contract and any other matters of strategic importance to the collection

Practice area 3: Quality assurance principles, standards and checks by data suppliers

Prices are initially validated at data entry on the hand held devices and then by Kantar TNS Head Office. They are then further validated by ONS and outliers (determined both in terms of price levels and price movements from the previous month) are checked by the ONS's staff. Depending on the outcome of the checks by the ONS's staff, these outliers are then either re-input into the computer system, queried with Kantar TNS or discarded. Where an outlier is queried, Kantar TNS must respond to ONS within three working days.

The price quotations from each location are checked individually by Kantar TNS for quality and those that do not conform to standard are queried with the retailer as necessary, or may be sent to the ONS with an indicator that identifies them as suspect. All suspect price quotations are resolved by Kantar TNS within three working days. Prices must be checked if:

- minimum or maximum price is out of a specified range
- the percentage change between months is greater than a specified threshold
- there is invalid use of indicator codes
- there is a change to the item description without the use of an indicator code

Practice area 4: Producers' quality assurance investigations and documentation

A further audit of prices is carried out every month by ONS staff (Field Auditors). They check around 70 items in a maximum of 12 locations per month. Both the location and items for checking are identified using random sampling. The Field Auditors check four measures of accuracy at this stage.

Zero and non-zero prices recorded as failures

The Field Auditors check that the prices collected are correct. If the price is different to that collected by the price collector, Field Auditors are instructed to find out what was the price on collection day. This usually means asking the retailer or provider if an item was on sale on the collection day and how much it would have cost.

Wrong items

The Field Auditors check to see that the correct items are priced, for example the item description may say 500 to 1000 grams. Pricing a packet weighing 454 grams would be a wrong item.

High description errors

The Field Auditors check to see that the descriptions for the items are clear and have been followed. The descriptions must not include prices and must make it obvious what item is being priced. The consumer price indices are based around price chains, which only work if the same item is priced. The description therefore needs to be accurate, informative and unique to that item to ensure continuity of pricing.

Non-comparable and comparable checks

If an item needs to be replaced there are two possibilities, that the new item is comparable (so the price chain is not broken) or it is not comparable so a new price chain needs to be started. The Field Auditors check that new items have an appropriate code to identify whether a new price chain needs to be started.

These four measures of accuracy are key performance indicators used by the ONS in managing Kantar TNS performance. The aggregate data collated from the Field Auditors are used to determine whether targets and tolerances have been achieved.

There are also key performance indicators around the timeliness of data delivery and coverage. Coverage is calculated as the number of quotes provided as a percentage of the maximum number of quotes that could be collected. Where the tolerance is not met for a key performance indicator then a service credit is payable by the contractor.

4. Valuation office agency

Practice Area 1: Operational context and admin data collection

Valuation Office Agency (VOA) Rent Officers compile and maintain a database of private market rents in England. 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.

The most significant factor determining how the data is collected is that there is a lack of any legal obligation 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 circa 500,000 rents since 2009.

The main statutory duties that shape the Rent Officer data requirements.

1. 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

2. 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").

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. Therefore 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 or distribution. These include checks against volumes by category and property a different geographical output levels.

Statistics or a full data set of housing benefit claims at medium super output area geographical level by property type or size is not available, therefore 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 data set. The collection methodology, which is more of a framework, has evolved to satisfy the Rent Officer's data requirements to fulfill their statutory duties.

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

Other data attributes are collected to support comparable valuations for housing benefit and Fair Rent purposes and to enrich the data set 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 six main collection methods used by Rent Officers.

Visits or phone calls to lettings agents or 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 which can be used to generate a report which details 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 3% 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 key influencers and potential influencers or advocates as possible with the aim of generating new contacts and maintaining VOA's profile within the private rental market community.

Rent Officers validate all data collected using the mentioned methods 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 and where practical they aim to replenish data before it reaches the end of its 12-month lifespan.

There are factors which 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 was 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 change 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), and 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.

Practice Area 2: Communication with data supplier partners

VOA has worked 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 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 Information and Analysis, 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) exist between the VOA and us documenting activities such as: data requirements, data transfer process, data protection. A delivery schedule is included within an Annex to this SLA which is reviewed on an annual basis; this schedule has always been met. Under this SLA, we require 3 months' notice prior to 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 on an annual basis to ensure the aims, benefits and terms of the SLA are mutually acceptable or require some adjustments. Regular meetings between VOA and us are timetabled to discuss performance and future plans. These meetings follow the following 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 VOA's "Peer Review Group" meeting which focuses on the quality assurance process for the development of VOA's own Private Rental Market Statistics. As a user of VOA statistics, we are also a member of VOA's "Domestic Statistics Advisory Panel" which meets on a bi-annual basis.

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

Practice Area 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 is a range of quality assurance and data validation processes in each collection, which complement those required by us, and take place either as the data is entered onto the system or as a part of the 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 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 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 one sole room (sole rooms are the product of sole bedrooms and sole living rooms)
- whole “house” or “flat” with shared bedroom, living room, kitchen or bathroom
- “studio” with sole rooms <> 1
- “rooms” (used for letting of 2 or more shared rooms in non-self-contained) with one sole room
- “room” (letting) with more than one sole room
- tenancy start date is before 15 January 1989, that is regulated tenancy – deleted centrally by CST
- dwelling type = unique
- property type = hostel
- “bedspace” with sole rooms greater than 0
- >25% variance between achieved and advertised rent (rents negotiated down or forced up by >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 Central Support Team (CST) is reported back to the responsible Rent Officer. Unsatisfactory responses are returned to the Rent Officer. CST retains, and will action, where issues are satisfactorily resolved. Around 3.7% on average goes on to be either amended (approx. 0.9%%) or deleted (approx. 2.8%).

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

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.

Information and Analysis request the data extracts directly from Digital 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 Service Level Agreement.

An illustration of this process can be found in “VOA data flow diagram – data collection” and “VOA data flow diagram – OOH processing” which can be found at figures 1 and 2 respectively of Annex C. A more generalised, high-level flow diagram of our processes for OOH data can be found at Figure 5 of Annex C.

Practice Area 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 four 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 Service Level Agreement, we require 3 months’ notice prior to 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 Annex B of the article '[Improvements to the measurement of owner occupiers housing costs and private housing rental prices](#)'. These thresholds will be reviewed on an annual basis as a longer diagnostics series becomes available.

Elementary aggregates

Elementary aggregate data for England (VOA) is combined with that of Scotland and Wales within the ONS system with this process being run by two 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 costs (OOH), Index of private housing rental prices ([IPHRP](#)), Consumer Price 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 Practice Area 2, are used to review the latest month and discuss any long-term trends in the data and its drivers.

Quality management

In March 2015 we commissioned an external audit of VOA processes. Six observations were identified to help with the future development of the procedure which has predominantly been implemented. The next external audit is scheduled for September 2016, these now continue on an annual basis.

Within ONS, 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. From September 2016 the processing of owner occupiers housing costs and private rental data (within ONS) become part of this quality management system.

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 valuation office Agency private rental market statistics, both of which are based on the same underlying data collected by VOA rent officers. We will [update this article](#) in October 2016 and look to extend this analysis during 2017.

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 QMI](#) (which uses the same underlying data) and [CPIH compendium](#).

5. Living Cost and Food Survey (LCF)

Practice area 1: Operational context and admin data collection

Living Cost and Food Survey (LCF) is a continuous survey collecting data on household expenditure on goods and services from a sample of around 5,000 responding households in Great Britain and Northern Ireland. The Northern Ireland survey is carried out by the Northern Ireland Statistics and Research Agency (NISRA). The LCF is a voluntary survey and involves a household questionnaire, an individual questionnaire and a detailed expenditure diary completed by each individual in the household over a period of 2 weeks.

Processes

There are five steps to the processing of the LCF survey:

1. Questionnaire/diary: information on regular expenditure is captured within Blaise, the software used to program the LCF questionnaire, during a face to face interview. Daily expenditure is collected in a two-week expenditure diary. This data is then coded into a Blaise questionnaire by ONS. Automated checks are built into both Blaise instruments.
2. Coding and editing: A summary of the coding and editing function is included in the LCF technical report.
3. Quarterly data processing: Derived variables are calculated from the Blaise questionnaires. These are created in the Manipula software, and these scripts are updated on a quarterly basis. Once the quarterly files have been run, the quarterly research checks are completed to ensure there are no errors in the datasets.
4. Annual data processing: Quarterly files are combined with reissues and imputation of partial cases is carried out. Research checks are repeated to ensure consistency. Expenditure outliers are identified by the LCF research team. The top five values for each COPCOP category are investigated. This process is currently under review.
5. Prices Delivery Process: LCF have provided a high level flow chart which describes the prices delivery process. The specifications are updated manually in excel templates to reflect in-year questionnaire changes; SAS scripts are updated manually to reference current year data sets, with the remainder of the process being automated. Checks are carried out using SPSS to provide an immediate flag of any errors in outputs.

Practice area 2: Communication with data supplier partners

The ONS employs a field force to deliver the questionnaires, and the LCF team hold regular communications with them. New interviewers are required to attend a briefing day, and supplied with instructions and background information about the survey prior to being assigned an LCF quota. They are also asked to complete an LCF diary for seven days, which is checked with feedback provided.

Refresher postal briefings are also available for interviewers who haven't completed any LCF quotas in the recent past. Annual questionnaire changes and other in year survey changes are also cascaded via a monthly newsletter. Interviewers all periodically contact the Research team to provide feedback.

LCF held focus groups with a subset of interviewers to gain feedback on the data collection process, Findings from the focus groups are summarized in [chapter 4 of the LCF NSQR report](#).

Users and uses

Historically the LCF was created to provide information on spending patterns for the Retail Prices index (RPI). It is now used for National and Regional Accounts to compile estimates of household final consumption expenditure. This is then used to calculate weights for Consumer Prices Index including owner occupiers' housing costs (CPIH), Consumer Price Index (CPI), and Purchasing Power Parities (PPP) for international price comparisons.

The Pay review bodies governing the salaries of HM Armed Forces and the medical and dental professions use LCF expenditure data.

Eurostat, and other government departments such as DECC, HMRC and Department for Transport also use the data.

Internally, National and Regional accounts use LCF data to compile estimates of household final consumption expenditure; they also provide weights for the Consumer Price Indices and for Purchasing Power Parities.

The LCF uses the expenditure classification system COICOP (classification of individual consumption by purpose), and has adopted international definitions and methods.

Practice area 3: Quality assurance principles, standards and checks by data suppliers

Quality assurance and validation checks

Systematic checks are applied to the aggregated data to ensure consistencies between diaries and interviews. Checks are also made to examine the processing of food types known to have problems with issues such as coding in the past.

Further checks are made by the LCF business operations and research staff. These include checks on missing shop codes, and checks on unit costs outside a pre-determined range. There are approximately 50 of these checks in total, and these are completed using SPSS. The checks can be categorised as follows with examples of individual checks provided:

1. High-level checks

- identify any missing values in key variables
- consistency in number of cases across datasets
- ensure imputation of certain variables has worked

2. Processing, editing and coding checks

- compares number of cases that editors and coders have completed with the number of cases on the delivered datasets
- further consistency checks – do all diaries have a corresponding person?

3. Questionnaire changes checks

Annual questionnaire changes are reflected in the microdata files, for example:

- derived variables reflect questionnaire changes

4. Validation and QA checks

- investigation of extreme values
- identification of items incorrectly coded

Further information can be found in Chapters 4 and 6 of the [LCF technical report](#).

In addition to the quarterly checks carried out on the LCF data as described above, each prices output is sense checked. Cells where data are above or below a certain percentage compared with the previous year are flagged for further investigation to understand the cause of the difference.

Missing data or imputation

LCF relies on households and individuals to complete their response. In 2015 to 2016, 170 households had imputed diaries, accounting for 3% of responding households. In the weighted data set, the imputed diaries accounted for 2% of people and 4% of households. A nearest neighbour hot deck imputation method is used for missing diaries, so the individual will receive the same data as another responding individual with matching characteristics of age, employment status and relationship to the household reference person.

There are several ways in which missing data can be imputed elsewhere. This can be done by reference to non-LCF data published elsewhere, for example imputing mortgage data based on tables containing interest rates and the amounts of a loan. Data can also be imputed by reference to average amounts from previous LCF data or by using information collected elsewhere in the questionnaire or by referring back to the interviewers.

Further information on imputation can be found in Chapter 5 of the [LCF NSQR report](#).

Review of processes

There are projects in place to ensure the continuous improvement of the LCF. For example, a project was implemented that examined the quality assurance process of the monthly data files produced for Defra, with the aim of reducing the resources required to complete the checking whilst maintaining quality.

In March 2016 the LCF National Statistics Quality review was published, which included an assessment of all aspects of the survey. We published a response to the NSQR.

The Prices delivery system was reviewed and rewritten in early 2014 to address concerns about the inefficiency of the previous processing and production systems which had the potential to lower the quality of the LCF statistics.

The current system is more robust, efficient and less error prone, given the reduction in manual intervention required. Data outputs were dual run for the period 2012 to 2013 to ensure consistency, moving to SAS only outputs for the 2013 to 2014 run.

Quality assurance checks are reviewed annually in consultation with the Prices team. Feedback is then incorporated into checking scripts ahead of generating the following years output.

Revisions policy

Provisional quarterly datasets are delivered to National Accounts 6 weeks after data collection is completed. Revised datasets are delivered alongside the following provisional quarterly delivery. Quarterly deliveries of data exclude partial and reissue cases- which represent a small proportion of each quarterly file.

During the process of finalising the financial year file, partial and reissue cases are incorporated.

National Accounts receive the finalised financial year file in October of each year. Incorporation of this data within the NA process is dependent on the blue book timetable which changes each year.

The Family Spending and HIE outputs are based on the full financial year files which includes all responding cases so no subsequent revisions are made to the financial year dataset.

No revisions are made to Prices outputs following the first delivery.

6. Mintel

Practice area 1: Operational context and admin data collection

Mintel publishes detailed descriptions of its data collection arrangements and operational context [on its website](#). When requested, they also produced a more comprehensive document detailing their data collection procedures, Quality Assurance methods and auditing practices. This document can be found at Appendix B.

Mintel constructs its reports using data from a variety of sources, including contracted agencies. Full details of these can be found in B, and illustrated in a flow diagram in Figure 7 of Annex C. They retain a large degree of control over this data by creating and quality checking surveys for these companies to use in the collection of data.

We have a contract with Mintel for 15 licences, which enables up to 15 members of prices division to log into the client pages of the Mintel website, and view their reports online.

The implications of accuracy and data are that weights for large portions of CPI would be incorrect. If access to website is restricted, we would have to source data from alternative sources.

Practice area 2: Communication with data supplier partners

The access to Mintel reports is decreed by a 2-year contract, which is a service level agreement that has clear specifications for data requirements and arrangements. This contract is renewed every 2 years, after being put out to tender.

The licence gives us access to the website, where the data transfer process consists of reports being viewed and downloaded. This contract was signed off by us and Mintel. The reports include a summary of key points as they relate to consumer behaviour for a product, graphs and tables summarising the data, and written descriptions for the results and the reasons behind trends.

There is a clear and established point of contact for Mintel, whose contact details are easily accessible via the website once a client has logged on. There are no regular meetings set, but ad-hoc meetings are established when needed. The point of contact for Mintel responds to email queries within a few days.

Mintel's point of contact has been quick to reply to communications and has produced all requested information promptly.

Practice area 3: Quality assurance principles, standards and checks by data suppliers

When requested as part of the Quality Assurance of Administrative Data (QAAD) assessment process, Mintel provided us with an in-depth document detailing their Quality Assurance processes and checks at all stages. (Annex A)

Mintel are full members of the UK Market Research Society and act in accordance to their guidelines.

Surveys are conducted by third-party companies – Lightspeed GMI for Online Surveys and Ipsos MORI for face-to-face surveys. These companies provide their own quality checks as well as being audited by Mintel. The surveys are designed by experts in Mintel's Consumer Research and Data Analytics team (CRDA), and are quality assured and signed off before being sent to the third parties.

Practice area 4: Producers' quality assurance investigations and documentation

Figures extracted from the latest reports are checked against last year's data for any obvious inconsistencies, such as figures that are significantly higher or lower than previous years. Confidence checks are also made against other data sources by searching for the product using a standard search engine. Once these checks are complete they are used in the construction of the CPIH.

Refer to Annex B for further details on producers QA checks

7. Scottish Government

7a.Private Rent Data

Practice area 1: Operational context and admin data collection

Rental data for Scotland which is currently provided by Rent Service Scotland (formally known as the Rent Registration Service), who are 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 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 five 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.

Practice area 2: Communication with data supplier partners

A Service Level Agreements (SLAs) exist between the Scottish government and ONS documenting activities such as: Data requirements, Data Transfer process, Data protection. A delivery schedule is included within an Annex to this SLA which is reviewed on an annual basis, this schedule has always been met. Under this SLA, ONS require 3 months' notice prior to any changes in data collection practices or format/coding of the data delivery to ensure sufficient time for amendments and testing.

Agreements are reviewed on an annual basis to ensure 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.

Practice area 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 LHA

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

Cross checks against known 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 (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.

Removal of duplicate

This is a two-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:

Zero bedrooms

- Non self-contained properties that are not "rooms"
- Number of living rooms exceeds number of bedrooms
- Property type "studio" with more than one room
- Self-contained "room" lettings
- Number of bedrooms exceeds seven

Rents listed as including gas and or electricity costs but without a value deducted from the rent for this component.

Audit

Hard copy data matching

Rent officers are responsible for quality assuring all lettings information (LI) entered onto 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 Rental 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 is scanned in later by the admin team.

ROCAS produces a report "Local Housing Allowance (LHA) with Letting information" which the Letting research 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.

Accompanied visits

During the course of the year the rental 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.

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 “Scottish government flow diagram- data collection”, which can be found following Practice Area 4.

Practice area 4: Producers’ quality assurance investigations and documentation

The Scottish government provides ONS with microdata on private rental properties which are loaded into the ONS 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. They, however, 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 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, however, 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 the ONS system with this process being run by two 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 which are based on the same underlying data (that is, comparisons are made between owner occupiers’ housing costs

(OOH), Index of private housing rental prices ([IPHRP](#)), Consumer Price Index (CPI) and Retail Price Index (RPI)). Any unexpected movements within the series are investigated using the raw data. If necessary these are queried with the data provider who can help aid by advising on perhaps regional policy changes.

Comparisons with other sources

Comparisons are made between the annual growth of IPHRP and that of some private rental providers. This comparison is 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 ONS private rental indices and Valuation Office Agency private rental market statistics, both of which are based on the same underlying data collected by VOA rent officers. We [updated this article](#) in October 2016 and look to extend this analysis during 2017.

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](#)' and the [Index of Private Housing rental prices QMI](#) (which uses the same underlying 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 will become part of this quality management system.

7b.Dwelling Stock Data

Practice area 1: Operational context and admin data collection

Scottish government dwelling stock count data are used to calculate strata weights, which are used to mix adjust the housing component of CPIH to reflect the OOH market. This is constructed using a mix of administrative data and survey data. Counts of all dwellings and vacant dwellings are taken from the National Records of Scotland's published figures, which are derived from council tax data. Local authorities provide Scottish government with counts of local authority social housing stock and vacant dwellings in statistics published in Housing Statistics for Scotland. Counts of housing association social housing stock and vacant dwellings are based on data collected and published by the Scottish Housing Regulator. Additionally the proportion of properties that are either owner occupied or privately rented is derived from Scottish Household Survey Data.

Processes

Information from the various sources are used to subtract the local authority and housing association stock, number of vacant dwellings and the number of privately rented properties to give an estimate of the number of owner occupied stock. This is done in March of each year. These stock by tenure estimates are published as national Statistics in Housing Statistics for Scotland, and following publication the DCLG Live Table 107 on Dwelling Stock in Scotland is updated.

The calculations are carried out in excel in a workbook which contains the necessary formulae. The workbook is updated manually when new figures are available. To mitigate the risk of human error, the figures are checked by a second person in the team, with the final figures cross checked to Scottish Household Survey results and census data.

Practice area 2: Communication with data supplier partners

The Scottish Government Housing Statistics team keep in regular contact with the National Records of Scotland's household estimates/projection team, Scottish Housing Regulator analysts, and Scottish Household Survey analysts, typically meeting with each on an annual basis or more frequently to discuss a variety of housing statistics matters. Communications with local authorities are generally through emails as part of the annual housing return collection and publication process.

Users and uses

Housing statistics for Scotland outputs are used as evidence for housing market analysts, forecasters and decision makers, feature in media reports on the housing market, and by academics across the UK. Local authorities used the statistics to plan their housing need and Demand assessments (HNDAs).

Stock by tenure estimates are used to help track changes in the sizes and proportions of each tenure category. Information on the sizes of each tenure category can be used for estimating the scale of properties involved in various government policy interventions.

Similar estimates of numbers and proportions of households by tenure are provided by the Scottish Government Household Survey Annual Reports. They only use the Survey data for these estimates along with all-household counts from the National Records of Scotland.

Practice area 3: Quality assurance principles, standards and checks by data suppliers

Quality assurance

Local authority data is quality assured by Scottish government prior to publication. These are compared with previous years; any discrepancies are queried with the local authority and, if necessary, changes are made either to the latest year's data or earlier to ensure consistency.

Tenure estimates are compared with full tenure results from the Scottish Housing Survey and figures from the population census. It has been found that all three sources are largely consistent with each other.

Figures from local authorities are compared to those in previous years to identify any discrepancies in the trends. Any discrepancies are queried with the local authority and if necessary changes are made either to the latest year's data or to earlier years to ensure consistency.

Internal quality assurance is undertaken on the stock by tenure calculations and estimates. The final estimates are compared with the Scottish Household Survey tenure estimates and with figures derived from the census.

Missing data or imputation

The main administrative datasets used are based on council tax data or social landlord stock data and so are established data sources that should have full coverage in relation to all-dwellings and all social sector dwellings. The Scottish Housing Regulator figures on vacant housing association stock relate to vacant normal letting stock only as opposed to all vacant stock (due to how this information is collected by the Regulator), that is any vacant stock that is not normal letting stock is excluded, such as stock awaiting demolition or modernization.

This only affects vacant stock figures for housing associations, as separate data collected direct from local authorities covers all vacant council house dwellings. However this is likely to have a very minor impact on final owner occupier figures, for example 1,000 extra vacant housing association properties, an estimate of the likely scale of magnitude based assuming that housing association vacant stock has the same profile as local authority vacant stock, would increase the calculated figure for occupied owner occupier dwellings by only around 1k dwellings out of 1,476k dwellings (0.05%). The calculated figure for occupied owner occupier dwelling would increase slightly given that an increase in the estimate for social vacant properties would reduce the estimate for private vacant properties, which is subtracted off the private sector total.

There is no imputation carried out when calculating the Scotland level stock by tenure estimates. For local authority area level estimates, some estimation is used for vacant housing association stock, where the Scotland level figure is apportioned out to each area based on previous data collected. However, this estimation is at the sub-Scottish level, and so does not impact on expenditure used by Prices.

Review of processes

There have been no reviews to the Housing Statistics for Scotland collection and publication since the statistics was reviewed for National Statistics status by the then UK Statistics Authority (now Office for Statistics Regulation) in 2012. There have been no fundamental changes to the methodology and data sources since then.

Revisions policy

Scottish Government provided a link to [the current published revisions policy](#).

There are two types of revisions that the policy covers.

Scheduled revision

Figures which are expected to be revised are clearly marked, along with an indication of the possible scale and nature of the revision if possible, and are incorporated in the next scheduled release.

Non-scheduled revisions

Minor errors will be corrected in the next edition of the publication, with the correction made clear and the reasons explained. Substantial errors are corrected on the website with the nature and extent of them made clear. Users are also notified of any errors which could affect their own work. Advance notice is given along with an expected release date and an indication of scale if it has been identified that the error will take time to correct.

Practice area 4: Producers' quality assurance investigations and documentation

For a general outline of QA procedures by us, which applies to Scottish Government, please refer to Annex B

8. Welsh government

8a. Private Rental Data

Practice area 1: Operational context and admin data collection

Rent Officers Wales which 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 is 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 five rent officers in the Market Rental Evidence Team and one line manager responsible for data collection within the 22 broad rental market areas (BRMAs) throughout Wales. There are seven main collection methods used by rent officers:

- Visits
- Letting lists
- Telephone
- Main 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 private rental sector (PRS). The most accurate data currently available is the Census 2011 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 a representative of the market as possible, however it is dependent upon the goodwill of agents and landlords for its provision. Landlords who only let one or two 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.

Practice area 2: Communication with data supplier partners

A Service Level Agreements (SLAs) exist between the Welsh government and us documenting activities such as: data requirements, data transfer process, data protection. A delivery schedule is included within an Annex to this SLA which is reviewed on an annual basis, this schedule has always been met. Under this SLA, We require 3 months' notice prior to any changes in data collection practices or format or coding of the data delivery to ensure sufficient time for amendments and testing.

Agreements are reviewed on an annual basis to ensure 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.

Practice area 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 Census 2011 so this is taken as the baseline for establishing the required sample.

Quality checks required for LHA

There is a range of quality assurance and data validation processes in each collection, which complement those required by us, and take place either as the data is entered onto the system or as a part of the 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 (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.

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 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:

- zero bedrooms
- non-self-contained properties that are not "rooms"

- number of living rooms exceeds number of bedrooms
- property type “Studio” with more than one room
- self-contained “room” lettings
- number of bedrooms exceeds seven
- rents listed as including gas and or electricity costs but without a value deducted from the rent for this component.

Audit

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.

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.

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 “Welsh government flow diagram-data collection”, which can be found following in Figure 3 of Annex C. A more generalised, high-level flow diagram of our processes for OOH data can be found at Figure 5 of Annex C.

Practice area 4: Producers’ quality assurance investigations and documentation

The Welsh 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; they, however, 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 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, however, 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 with this process being run by two 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 which are based on the same underlying data (that is comparisons are made between owner occupiers' housing costs (OOH), Index of private housing rental prices ([IPHRP](#)), Consumer Price 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 aid by advising on perhaps regional policy changes.

Comparisons with other sources

Comparisons are made between the annual growth of IPHRP and that of some private rental providers. This comparison is 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 Valuation Office Agency private rental market statistics, both of which are based on the same underlying data collected by VOA rent officers.

This article was [updated](#) in October 2016 and we will be extending this analysis during 2017.

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](#)' and the [Index of Private Housing rental prices QMI](#) (which uses the same underlying 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 will become part of this quality management system.

8b. Dwelling Stock Data

Practice Area 1: Operational context and data collection

Welsh Government dwelling stock count data are used in the production of CPIH. The release draws on information from a range of data sources in order to compile a coherent set of statistics on the total number of dwellings and the tenure profile of the stock. The sources include, but are not limited to, census data from 2011 and 2001, the annual population survey from the Office for National Statistics, and local authority stock and registered social landlord stock from Welsh government.

Estimates of the total dwelling stock are calculated based on data from the population censuses. The estimates shown in this release are produced by using the dwelling count from the most recent 2011 census as a baseline. This count is then projected forward using information collected on annual changes to the dwelling stock through new build completions plus any gains or losses through conversions and demolitions.

Further information on the [differences between the 2001 and 2011 Census](#) is available in a series of evaluation reports produced by the Office of National Statistics.

The breakdown of stock estimates by tenure shown in this release is estimated from 2011 Census information, information from the Annual Population Survey, local authority returns and registered social landlord (RSL) returns. This information takes into account any changes in tenure through sales and acquisitions.

Social sector housing – local authority and registered social landlord dwellings

1. The data on local authority and registered social landlord housing stock are taken from the annual returns from social landlords and is available on the Stats Wales interactive website.
2. This data is used directly in the dwelling stock tenure split and include all self-contained and non self-contained dwellings but exclude intermediate and other tenures which are not at social rents, which are included in the owner-occupied, privately rented and other tenures category. The data excludes all non-residential properties and excludes any dwellings leased to temporarily house the homeless and any dwellings that are managed as a social lettings agency.
3. As the annual returns collect the number of non self-contained bed spaces rather than dwellings, it is assumed on average that three non self-contained bed spaces is equal to one dwelling. Information on the number of non self-contained units for intermediate and other tenures is not collected; therefore the same calculation cannot be applied for these tenures.

Private sector dwellings

4. Private sector dwellings are calculated by subtracting the number of local authority dwellings and RSL dwellings from the total number of dwellings in Wales.
5. Whilst private sector stock covers both owner-occupied¹ and private rented dwellings, there is no direct measure of these tenures due to the difficulty of collecting information on the private sector and the relatively fluid interchange between these two parts of the private dwelling stock.

Owner occupied and private rented dwellings

6. In order to estimate the number of private sector dwellings that are privately rented the current methodology estimates what proportion of the private sector are privately rented using information from the Annual Population Survey (APS). The owner-occupied tenure is then calculated as the residual after the other tenures have been removed.
7. The APS is a boosted version of the Labour Force Survey (LFS). Like the LFS the APS provides estimates for the private rental sector but it only covers occupied dwellings, therefore no account is taken of vacancy rates in producing the split. Unlike the LFS, the APS is based on a sufficiently large enough sample to provide a separate

percentage breakdown for privately rented stock at a local authority level within Wales. For 2015 to 2016 the percentage of private rented dwellings at an individual local authority level has been calculated using the information from the APS.

8. The APS is a survey of households living at private addresses in the UK (therefore NHS accommodation, prisons and army barracks are excluded). The purpose of the LFS is to provide the information on the UK labour market required by the European Statistical Office (EuroStat) under the Treaty of Rome. The APS is boosted by the Welsh Government to collect a wide variety of information from labour market situation to education, health, place of residence and work and household and family characteristics.

Practice Area 2: Communications with data supplier

Several different communications are undertaken with supplier; these include in-form validation queries, which seek clarification on data provided where it does not meet pre-built in validation requirements, and out of form validation queries where data falls outside of an expected variance compared to other providers.

There are also specific work groups for data projects; for instance the social housing rent standard where data was discussed to improve clarification to guidance.

Users and uses

The dwelling stock estimates are used as evidence in policy making by both central and local government. The information provides an estimate of the number of residential dwellings by each tenure type and by local authority, at the end of March each year. The data are used by the Welsh government, local authorities and other housing organisations to help monitor trends in the overall level of Welsh housing stock, as well as any changes in its tenure distribution over time.

The dwelling stock estimates provide annual base line information on the overall amount of housing stock at a Wales and local authority level. It is used as evidence for policy making by both central and local government. The data are used by the Welsh government, local authorities and other housing organisations to help monitor trends in the overall level of Welsh housing stock, as well as any changes in its tenure distribution over time. The data are also used by the Welsh government in the calculation of local government standard spending assessments.

Local authorities use dwelling stock information to develop their Local Housing Market Assessments; for benchmarking; for evidencing how housing need and demand is being met locally and for assessing future requirement and need in order to plan and allocate resources effectively. Outside of government the dwelling stock estimates are used by the finance and investment industries, for example to help develop a picture of demographic trends.

The statistics have a number of uses, for example:

- advice to Ministers
- to measure government targets and key performance indicators
- to provide context and evidence for the Welsh Government's National Housing Strategy
- unitary authority comparisons and benchmarking
- to compare housing in Wales to other countries
- to inform the debate in the National Assembly for Wales and beyond
- to assist in housing research and analysis
- housing revenue account subsidy and other housing finance calculations
- local government finance standard spending assessment calculations
- compendia publications by other organisations (for example, Regional Trends produced by ONS, Welsh Housing Review by the Chartered Institute of Housing, and UK Housing Review)

It is believed that the key users of housing statistics are:

- ministers
- Assembly Members and the Members Research Service in the National Assembly for Wales
- local government unitary authorities (elected members and officials)
- National Park authorities
- registered social landlords
- Welsh Local Government Association
- Community Housing Cymru
- Her Majesty's Treasury
- the Office for National Statistics
- Department for Communities and Local Government
- Chartered Institute of Housing
- Shelter Cymru
- Chartered Institute of Public Finance and Accountancy
- students, academics and universities
- other colleagues within the Welsh government
- other government departments
- individual citizens and private companies

Practice area 3: Quality assurance principles, standards and checks by data suppliers

Data are collected from local authorities via Excel spreadsheets. These are downloaded from the Afon file transfer website which provides a secure method for users to submit data.

The spreadsheets allow respondents to validate some data before sending to the Welsh government. Respondents are also given an opportunity to include contextual information where large changes have occurred (for example, data items changing by more than 10% compared to the previous year). This enables some data cleansing at source and minimizes follow-up queries.

Local authorities are notified of the data collection exercise timetable in advance. This allows adequate time for local authorities to collate their information, and to raise any issues they may have. There is guidance in the spreadsheet, which assists users on completing the form.

Examples of validation checks within the forms include year-on-year changes, cross checks with other relevant data tables and checks to ensure data is logically consistent.

Once the data is received, it goes through further validation and verification checks, for example:

- common sense check for any missing or incorrect data without any explanation
- arithmetic consistency checks
- cross checks against the data for the previous year
- cross checks with other relevant data collections
- thorough tolerance checks

- verification that data outside of tolerances are actually correct

If there is a validation error, the organization is contacted to seek resolution. If an answer not received within a reasonable timescale, imputation is then used to fix the error. The organisation is then informed and it is explained to them how the data has been amended or imputed. The method of imputation and the affected data is highlighted in the “quality information” section of the first release.

All data collected is loaded to an SQL database directly from a data loading sheet within each data providers return. Each data item and all calculated totaling is checked via a secondary independent process, ensuring all data provided and loaded to the database is accurate. The data from the database populates both release tables and data published to StatsWales.

The release is independently checked and a final sense check is carried out by the Housing Statistician prior to publication on the website.

Statswales data is further checked against data provider returns as well as the first release.

Revisions

The data shown in the quarterly and annual releases are final at the point of publication. Following publication revisions to the data can arise from events such as late returns from a local authority or when a data supplier notifies the Welsh government that they have submitted incorrect information and resubmits this. Occasionally, revisions can occur due to errors in our statistical processes. In both these cases, a judgement is made as to whether the change is significant enough to publish a revised statistical release. Significant revisions to the data will be addressed with a revised release and users informed in accordance with the Welsh Government’s Revisions, Errors and Postponements arrangements. Where revisions are not deemed to be significant, that is, minor amendments, these will be reflected in the StatsWales tables and in the next version of this release. However minor amendments to the figures may be reflected in the StatsWales tables prior to the publication of that next release.

The estimate of total dwellings in the 2011 Census was higher than the rolled forward estimate for 31 March 2011 by 34,178 dwellings. To ensure consistency with the 2011 Census figures the dwelling stock estimates for Wales and the individual local authorities from 2001 to 2002 to 2010 to 2011 were revised based on the 2011 Census figures.

Practice area 4: Producers’ quality assurance investigations and documentation

9. Department for Business, Energy and Industrial Strategy (BEIS)

Practice area 1: Operational context and admin data collection

Data for Road Fuel Price Statistics Bulletin, produced by the Department for Business, Energy and Industrial Strategy (BEIS) are based on weekly and monthly surveys. Six companies (four oil companies and two supermarkets) are surveyed as part of the weekly fuel price survey, providing ULSP (unleaded petrol), ULSD (Diesel) and super unleaded fuel prices. These cover around 65% of the market. The fuel companies are contacted by email every Monday morning to gather their fuel prices for that day.

The survey is administered by BEIS staff who receive survey returns via email. In addition to the above companies, every month one extra oil company and two extra supermarkets are contacted by email. The response rate is excellent with regard to the data collection on the road fuel prices. Suppliers have been complying as expected, despite surveys currently being voluntary. On the rare occasion when it is not possible to contact a company, an estimated value will be calculated for that company. In general, prices follow a similar pattern so the average price change will normally be estimated based on a paired company.

Data are split into two strata (supermarket and other) and are weighted separately to reflect the whole market.

Processing

Prices are entered manually onto a spreadsheet in order to calculate the weighted price for each fuel and then averaged to produce the weekly price.

The data published are national average prices calculated from prices supplied by all major motor fuel marketing companies. Sales by super or hyper markets are also included in the price estimates.

Practice area 2: Communication with data supplier partners

Users and uses

Road fuel price data are collected to meet EU Commission requirements (Council Decision 6268/99) and to publish on their website. Data also made available on BEIS website as National Statistics and re-published by motor organisations (such as the RAC), and by ONS for CPIH and Consumer Price Index (CPI). Data are also supplied to the Bank of England and other commentators. Road fuel prices published by BEIS are UK National Statistics.

Practice area 3: Quality assurance principles, standards and checks by data suppliers

Quality assurance

For road fuel prices, there are a number of quality assurance checks in place, for example:

- look at the trends in the data
- check how individual suppliers fare against each other in the same category (oil and supermarket); for example does the high price supplier tend to remain above the others all the time, given the price fluctuation? Are the supermarkets consistently low?
- compare trends with other data sources; for example, Experian data
- compare prices with wholesale oil (brent crude); there is normally a 6 week or so lag in change in prices of crude oil feeding through to retail petrol prices
- checking if the results are in line with press stories on price cuts and rises

Sense checks are also carried out on the final outputs to identify further errors.

For more information, please see the [Domestic Energy Prices Methodology document](#).

Practice area 4: Producers' quality assurance investigations and documentation

For a general outline of QA procedures by us, which applies to Department for Transport, please refer to Annex B

10. Brochures, reports and bulletins

Background to data

Some individual prices and expenditure details are accessed via brochures, reports or other non-statistical bulletins. For instance the latest edition of GB Tourist is used to calculate the weights of UK Holidays (excluding self catering). Similarly expenditure for newspapers and periodicals below the item level is derived from ABS National Newspaper reports of monthly average net circulations.

This QAAD assessment will encompass all sources of this type, and will assess the procedures used to identify a source, choose it for inclusion over alternatives, incorporate the information into the CPIH calculations, and ensure the information obtained is as accurate and relevant as possible.

Practice area 1: Operational context and admin data collection

Brochures and other hard media are generally used for year-on-year comparisons, and so are purchased annually.

They are purchased on subscription, and delivered to our prices production team.

The current brochures were chosen several years ago, and have been consistently used in the Consumer Price Index (CPI) since.

Practice area 2: Communication with data supplier partners

Each publication has a listed contact which the prices production team can contact with any queries. The publications arrive regularly by post. There have been no reported delays in receiving the reports, which have always arrived on time for the data to be included in the index.

Practice area 3: Quality assurance principles, standards and checks by data suppliers

Publications such as GB Tourist and ABC follow their own procedures for data collection and quality assurance. As this is a general assessment details for individual publications shall not be provided, however all sources have a corresponding website which contains information on their practices.

Practice area 4: Producers' quality assurance investigations and documentation

Prices Production areas are externally accredited under the quality standard ISO9001, which promotes the adoption of a process approach, which will enable understanding and consistency in meeting requirements, considering processes in terms of added value, effective process performance and improvements to processes based on evidence and information. These standards are adhered to when collecting from brochures and hard media.

Please see Annex B for a general list of procedures for inputting figures into the CPI

11. Consumer Intelligence

Practice area 1: Operational context and admin data collection

Quotes are supplied by consumer intelligence.

The weights are calculated from the market share of each insurance company. These shares are then rescaled as a percentage to form the company weights. The source for the market share figures is the Financial Services Authority (FSA) or the Association of British Insurers (ABI). The data can be requested from the FSA but the ABI source (also used for the car insurance) appears to be more reliable. The weights are derived for each individual quote by taking the company share and dividing this by the number of quotes for that company. The weights data is lagged by a year so, for example, the 2013 spreadsheets are based on 2011 data.

For 2013 onwards, the weights data for RBS combined the companies within the RBS Group – RBS, Direct Line and Churchill. To include these companies within the collection, we applied the 2010 weights data (used in the 2012 spreadsheet) for each company's market share (within the RBS Group).

Practice area 2: Communication with data supplier partners

There is an account director who is the main point of contact for enquiries. There is also an alternative contact available. Each month the contact sends through quotes for all UK dwelling insurance providers.

Practice area 3: Quality assurance principles, standards and checks by data suppliers

Due to issues with contacting the account director there is little information available on the Quality processes carried out by Consumer Intelligence. The information received is a selection of quotes from insurance providers, so there may be little processing involved before they are sent to us.

Practice area 4: Producers' quality assurance investigations and documentation

For a general outline of QA procedures by us, which applies to Consumer Intelligence, please refer to Annex B

12. Kantar

Practice area 1: Operational context and admin data collection

Kantar collect prices data through a consumer panel of 15,000 individuals. These are in the age range 13 to 59, and are stratified into Age, Gender and Region.

The consumer panel excludes Northern Ireland.

There is no contract or SLA in place for Kantar data. We purchase the data annually in a one-off payment.

Practice area 2: Communication with data supplier partners

There is a dedicated contact for any issues. When contacted, the Kantar representative agreed to a short telephone meeting to discuss their QA procedures. This was very productive and the representative provided thorough answers to the questions provided.

Practice area 3: Quality assurance principles, standards and checks by data suppliers

The classifications for stratification are matched with the stratifications used in our Census publication.

Individuals are given log on details to a system where they record information about their entertainment purchases. Every 4 weeks this data is collected.

The weights data comes from expenditure recorded by EPOS, who provide the electronic point of sale technology to retailers. This data is collected by third party companies and purchased by Kantar.

Around 80% to 90% of retailers are picked-up by this data collection. A notable exception is Toys-r-us, who do not provide data. Therefore if there is a product that is exclusively stocked by this retailer, then it will not be represented.

This data is monitored over time. Once the data has been collected, there are several Quality Assurance Systems and processes in place.

Practice area 4: Producers' quality assurance investigations and documentation

For a general outline of QA procedures by us, which applies to Kantar, please refer to Annex B.

13. Department for Transport

Practice area 1: Operational context and admin data collection

There are three elements to the Department for Transport (DfT) data which is used by Prices.

Rail fares

The data sources for the rail fares index are the LENNON ticketing and revenue database (admin data), a fares data feed from the Rail Delivery Group (admin data) and data pulled from the Retail Prices Index for comparative purposes. The LENNON ticketing and revenue database is used to source both weights information (all revenue from the year preceding the January price change) and ticket price data. Rail Delivery Group (RDG) now have a fares data feed available for download which provides ticket price info for all flows which we now use to match the weights data to the price data. The Retail Prices Index is used to compare price change in rail fares with price change in other goods and services.

DfT have provided a detailed outline of the processes used to produce the dataset. This is a combination of manual extraction and cleaning, and automatic processes in SPSS.

Light Rail and Tram

Tables [LRT0301a](#) and [LRT9902a](#) are constructed from the annual DfT Light Rail and Tram survey.

These are then compiled into a single excel spreadsheet, and manipulated into the format of the published tables.

Channel Tunnel data

DfT publish figures [for passengers, vehicles and freight trains using the channel tunnel rail link](#) in the a table.

Data sources

Vehicles carried on Le Shuttle and Eurostar passenger numbers are [sourced from published data](#) on the Eurotunnel Group website.

Unrounded numbers of 'passenger equivalents' for Le Shuttle are obtained directly from contacts at the ORR.

Unrounded figures for through-train freight tonnes are [sourced from an annual press notice](#).

Data published by Eurotunnel Group are input into a spreadsheet along with data sourced from the ORR. A simple summation is carried out on the disaggregated data provided by ORR (for Le Shuttle passengers) and the published Eurostar passenger numbers to produce a single number for channel tunnel passengers.

Practice area 2: Communication with data supplier partners

Rail fares

Quarterly bilateral meetings are held with the Rail Delivery Group (RDG) but they are not responsible for the supply of data, with the fares feed data being accessed through a login on the RDG website. However, they do update DfT if there are any expected delays to the data being uploaded to the website. DfT are in email contact with the Lennon support desk who provide any advice. There are no face-to-face meetings with them, apart from when they host workshops on developments within the LENNON system.

Government uses the data to inform ministerial briefings, to help set future policy and for inclusion in other government produced reports.

Media use the data to publish news articles and commentate on changes in rail fares.

Academia and consultants use the data as part of research projects.

Light rail and tram

No regular communications are held between the light rail and tram operators and the DfT statisticians.

The figures are used by:

- DfT – to inform briefings and to answer PQs
- academics – for teaching and research purposes
- industry – to provide insight into the effectiveness and impact of LRT systems, making comparisons between areas, over time, and with other modes of transport

Channel Tunnel data

As information is sourced directly from published material, there are no regular communications with the Eurotunnel Group.

Some unrounded figures are sourced from the ORR where the information is held to a higher level of accuracy, and there is engagement with ORR to obtain the information when required.

The statistics have been used by internal analysts and policy teams looking at EU exit and UK trade. Further underlying data, which includes 'direction of travel' information, has also been obtained from ORR for these purposes. There are also occasional public enquiries where these statistics may be used in a response.

Practice area 3: Quality assurance principles, standards and checks by data suppliers

Rail fares

Dft conducts a number of validation and quality assurance checks on the data, including:

1. carrying out checks on flow/product combinations where the price change is deemed unrealistic (generally outside the negative 20% or positive 20% price change range)
2. monitoring regulated price changes against the government price cap
3. checking TOC, sector or product price changes to pinpoint any irregularities and rectify

Data tables quality assured by a member of the Business Intelligence Team, looking at

1. Re-calculating the average price change from the indices provided
2. Re-calculating the real terms changes in average price from the Retail Price Index (RPI) figures
3. Check on the magnitude of the price changes
4. Note any revisions and ensure these are flagged appropriately

Statistical release reviewed by either the Head of Profession or Deputy Head of Profession

Light rail and tram

All DfT statistical publications have recently undergone an independent review, including the light rail and tram publication.

The figures published in LRT0301 are National Statistics. Light rail and tram statistics were assessed by the UK Statistics Authority and confirmed as National Statistics in February 2013.

The figures published in LRT9902a are outside the scope of National Statistics but are included to provide wider context.

All light rail and tram operators complete the survey; there is therefore no missing data.

Channel Tunnel

The DfT channel tunnel statistics are dependent on Eurotunnel Group publishing accurate and consistent information each year, and Dft do not have any direct involvement in their data collection and data quality processes.

No internal quality assurance measures are undertaken on the published source data. Data that are obtained from ORR are checked against published numbers that is to check the number rounds to the published number.

The DfT [statistics revisions policy is published online](#).

Practice area 4: Producers' quality assurance investigations and documentation

Rail fares

Heathrow Express fares information are not captured as they do not record their data within the Lennon database. The revenue for Heathrow Express is not known so DfT cannot be sure what percentage is not included but they do account for 0.35% of journeys so the assumption is that it would be a similar figure for revenue.

Furthermore, with the exception of advance fares, the index is constructed based on matched prices (that is the flow or ticket combination has a fare price in both reference years (Jan 2016 and Jan 2017)). Where there is no value in either of the two periods, these flows are excluded from the calculation of the percentage change. The flows that do get excluded tend to be very low revenue flows so although they are quite large in number (original dataset contains around 25million records, final index is calculated from around 3million records) in terms of their relative impact on the index, it is very low.

The volume of revenue used in the file to calculate the price changes is around 90% of the total revenue. However, following advice from the ONS Methodology the weight from the flows that have been excluded along the way are included in the final aggregation of data.

The process for calculating the rail fares index was reviewed by the ONS Methodology Advisory Service in 2013 to 2014. The mapping is reviewed each year; this includes updating the register of regulated or unregulated fares, checking that particular product codes are still being mapped to the correct categories (for example, advance).

Light rail and tram

The latest returns are compared to previous years. Any unexplained changes are followed up with the operator.

Channel Tunnel

Outputs are checked against source data by two members of the production team.

14. Direct contact

Background to data

Some price information in the Consumer Price Index (CPI) is collected by contacting the supplier of the item directly. This may take the form of a phone call to establish the cost of a service, for instance a hairdresser, or emailing a company to find out their price.

These types of price collection have been grouped together under Direct contact, which has undergone a general quality assurance (QA) assessment.

Practice area 1: Operational context and admin data collection

Data is collected as individual prices, however when the prices are input into the Pretium system, automatic error checking is applied, due to the system being designed for datasets. This is disregarded for direct contact as each price is individually checked.

Practice area 2: Communication with data supplier partners

A sample of small business and individuals are contacted monthly to determine if there has been any price change to their service or product. There is therefore regular communication with the supplier.

When ringing suppliers, prices staff are instructed not to quote the previous price of the service or product before being told the new price.

Practice area 3: Quality assurance principles, standards and checks by data suppliers

Direct contact involves ringing service providers for a quote. As such, there is little in the way of quality assurance that could be provided for this assessment other than the small business or individual ensuring that they give the correct price to the ONS employee when contacted.

Practice area 4: Producers' quality assurance investigations and documentation

Prices Production areas are externally accredited under the quality standard ISO9001, which promotes the adoption of a process approach, which will enable understanding and consistency in meeting requirements, considering processes in terms of added value, effective process performance and improvements to processes based on evidence and information. These standards are adhered to when collecting from direct contact sources.

Please see Annex B for a general list of procedures for inputting figures into the CPI.

15. Glasses

Practice Area 1: Operational context and admin data collection

Glasses receives their data from several sources. The National Association of Motor Auctions (NAMA), around 650,000 retail observations and through web portals such as motors.co.uk and AA Cars.

Meetings are also held with customers, motor trade experts, manufacturers, dealers and auctioneers.

We use the Glasses database to track 90 cars throughout the year. Until April 2016 this data was sent to our prices division in the form of a CD. As of April 2016, the data is accessed via the Glasses website, and we are provided with login details for the secure part of the website.

Since the switch from CDs to website access for glasses, members of Prices have reported extra difficulties in accessing and processing the data, in particular the use of codes.

It was reported that no training was provided by Glasses to ease the transition from CD to website. Members of the prices have indicated that the data is still fit for purpose, but additional time resources are currently required to find the correct values. This is anticipated to decrease in time as the team becomes used to the website.

It has also been noted however, that the move from CD to website has reduced some of the risk of the CD submissions, which were problematic and took up additional resources due to a piece of software required to extract the data not being supported.

Practice area 2: Communication with data supplier partners

There is an official contact for Prices as part of the contract. However they are not normally contacted as most requests are deemed to be more straightforward, such as asking why a car price has changed so much. For these requests there is a helpdesk that can be contacted by phone or email.

When conducting the assessment, it was found that the primary contact had not worked for the company for several years. Glasses provided an alternative contact where the QA questions could be sent.

Practice area 3: Quality assurance principles, standards and checks by data suppliers

Glasses publish an overview of their data processing procedures. This includes the various sources of their car price valuations. Their valuation process, which is essentially their quality assurance procedures, and their measures of Accuracy.

Also included is a comparison of accuracy, which is the Glasses Trade value as a percentage of the observed hammer price. There is also a comparison with their main rivals. These figures are released monthly, and are archived and available from July 2015.

This information is illustrated in Figure 8 of Annex C.

This information is available online, and although there is no reason to question its integrity, it should be noted that the document forms part of their marketing strategy to potential purchasers of their product.

Glasses have been contacted and asked to provide more in-depth information on their quality assurance procedures. A representative has stated that they can provide this information, however at the time of publication of this document it has not been received.

Practice area 4: Producers' quality assurance investigations and documentation

Processing the data

Once the data has been extracted from the Glass's Guide, it is checked by a prices analyst. Any queries at this point are raised with the Glasses helpdesk.

For the manual transfer method the data for 1, 2 and 3-year-old cars is input into the spreadsheet by the prices analyst making sure to input the prices into the worksheet for the appropriate year and to match the prices with the row and column titles.

The spreadsheet automatically calculates the overall indices. Spreadsheets are formatted so that yellow cells indicate data entry pink cells indicate the final indices. Blue text indicates an increase in the price compared with the previous month and red text indicates a decrease. Black indicates no change.

Checking the data

The spreadsheet is printed out by the prices analyst and passed to the checker together with the printout from Glass's Guide.

The checker must check that prices have been obtained for cars with the appropriate registration number. This is listed in column F of the spreadsheets. The mileage of the cars priced should also be checked. It should be 1,000 miles higher than in the previous month. In practice the average mileage quoted by Glass's Guide is used. Since April 2016 this can only be done by the checker logging into the Glass' website using the ONS log-in details found in the "obtaining the data" section of this document.

When the checker is satisfied that prices have been selected for cars of the correct age and mileage a check should be made that the prices have been correctly transcribed into the spreadsheet.

Once the spreadsheet has been checked, the indices are input onto the mainframe by the prices analyst¹ and the signed printout of the spreadsheet is put on the working file with the Price Data.

16. Higher Education Statistics Agency (HESA)

Practice area 1: Operational context and admin data collection

The data on the number of non-EU students attending each university is sent to Prices division in the form of an Excel Spreadsheet.

Practice area 2: Communication with data supplier partners

Prices Division has a Data Sharing agreement with the Department for Business, Energy & Industrial Strategy (BEIS) (formally known as the Department for Business, Innovation and Skills (BIS)) for Higher Education Statistics Agency (HESA) data. The contact for Prices had recently left BIS, and after contacting the department an alternative contact was proposed, although it was not clear whether they would be in the right position to help. The contact agreed to attempt to answer the questions relating to their quality assurance procedures.

The data is delivered regularly to Prices Division with no delays.

Practice area 3: Quality assurance principles, standards and checks by data suppliers

The HESA Student number publication is used to calculate the number of International Students. This is published on the HESA website, along with its quality Assurance procedures.

For Student number data, HESA provide a statement of administrative sources which covers data from Higher Education Institutes, plus 1 private institute the University Of Buckingham.

[HESA Statement of Administrative Sources](#)

Also provided was a [link to their website which contains quality assurance practices and data tables](#).

For student fee information, The Student Loans Company (SLC) and The Office for Fair Access (OFF) publish information on their publications and data quality.

[SLC Data Quality Statement](#)

[SLC Publications OFFA publication](#)

BIS projections are based on simple projections of Inflation

Practice area 4: Producers' quality assurance investigations and documentation

Once the data is inserted into a local spreadsheet, any unusual price movements, reasons for change, or other points of interest, are included in a "Data Notes" text box for future reference. The spreadsheet is set up to automatically calculate the index for CPIH once the data has been inputted.

Once complete, the spreadsheet is printed and checked by a member of prices division. Once these checks have been completed, the index is inputted into the main CPI index.

17. Home and Communities Agency

Practice Area 1: Operational context and data collection

The Homes and Communities Agency (HCA) uses rental price data for registered social landlords (RSL) for use in the production of CPIH. A statistics data return (SDR) is completed by private registered providers of social housing, via the online portal NROSH+.

SDR returns are stored securely with the NROSH+ infrastructure, accessible to the submitting PRP and HCA regulation staff. The individual returns are collated into a single data transfer file and are held within a restricted area on the HCA internal server.

The data transfer file is an excel file and is subject to checks to ensure consistency with the underlying data. These include pot checks to ensure individual Private Registered Provider (PRP) returns are captured correctly. Data submitted by providers is redacted within the public release to remove all contact information submitted within the Entity Level Information (ELI) section. This contact information is not publically available.

Practice Area 2: Communication with data suppliers

An annual letter is sent in March to CEOs of all providers informing them of the data collection requirements for the year ahead. The NROSH+ website, through which data is returned by providers, is also used to send emails and publish news articles, which are intended to remind providers of requirements and deadlines. A helpdesk is also available to providers should they require advice on completing the SDR, and a range of guidance materials and FAQs are provided on NROSH+.

Users and uses

HCA provided a user feedback document which contained the results of a survey.

The primary use of these statistics is for regulatory purposes, to determine sector characteristics and used as a basis on which to predict the impact of risk.

Practice Area 3: Quality assurance principles, standards and checks by data suppliers

HCA regulation (data team) subject submitted SDR data to a series of internal checks to identify potential quality issues before each individual data return is signed off. The final SDR data file that supports the statistical release is only created from individual SDR returns that are checked and signed off. Where outstanding queries, deemed material to the final data set, cannot be resolved, the data is excluded from the final data set. In 201/1 no returns were excluded.

There are two types of checks on SDR data submitted by providers: Automated validations, and manual inspection and sense checking.

Automated validations are programmed into the NROSH+ system and check the data at the point of submission. Checks include:

- ensuring every data point is in the correct format
- confirming whether data is consistent, logically possible and within expected limits
- automated validations are either “hard” or “soft”

“Hard” validations result in data that cannot be submitted by the providers without the issue being addressed. ‘Soft’ validations trigger a warning to the provider to check their data before submission.

Following submission and automated checks, the data team run a systematic programme of manual inspections and sense checking on all submitted data before it is signed off within NROSH+. Random spot checks on 10% of returns are also undertaken to ensure that the testing regime is robust.

For all providers with 1,000 or more units there is a full manual check of the data. New providers, those with affordable rent stock, or a degree of complexity in group structure or geographical stock ownership, are subject to further manual checks. Stand-alone PRPs with fewer than 1,000 units operating in a single local authority are subject to a basic check.

All returns are subject to tests which ensure changes in current and prior year stock totals are broadly consistent with submitted data on stock movement within the year and that reported group structures are consistent with other provider returns.

Where potential anomalies are detected with submitted data, a query is raised with the provider. The sign of returns for all providers with 1000 or more units is dependent on the resolution of all queries. Once a final data set is created no further amendments to the returns are possible. In 2015/2016 all queries were resolved with large providers.

Almost all data submitted by providers is published at a disaggregated level as part of the statistical release. Releasing data into the public domain serves as an additional route through which erroneous data may be identified by the provider or third parties.

Missing data or imputation

All providers are required to complete the SDR. Nevertheless due to either non-submission or exclusion due to unresolved errors there is still a level of non-response. In 2016 the overall non-response rate was 5%

Review of processes

Quality assurance checks are reviewed annually.

System validations and checks are reviewed during the development of the survey (September to October each year).

Manual checks on incoming data are reviewed and agreed during January to March each year.

Analysis QA procedures are reviewed and agreed during March to May each year.

Following the collection cycle, lessons learnt are captured and feed into the following year’s processes.

Revisions policy

Where producers report errors on data already submitted, these are recorded and used to correct data either in the subsequent years' statistical release, or through a supplementary release during the year if the level of error is deemed material to the use of the data. The level of revision due to identified errors is documented within the following year's statistical release.

If it is deemed that significant material errors have been submitted by the provider maybe that reasonably ought to have been found in the provider's quality control processes, then the regular will consider whether this offers evidence of failure to meet requirements for data quality and timeliness under the Governance and Financial Viability Standards. The most appropriate proportionate response will then be taken, taking into consideration data quality and timeliness issues across other regulatory data returns.

Practice Area 4: producers' quality assurance investigations and documentation

For a general outline of QA procedures by us, which applies to HCA, please refer to Annex B.

18. Inter-Departmental Business Register (IDBR)

Practice area 1: Operational context and admin data collection

The Inter-Departmental Business Register (IDBR) covers over 2.6 million businesses in all sectors of the UK economy, other than some very small businesses (those without employees and with turnover below the tax threshold) and some non-profit making organizations. The IDBR was introduced in 1994, and is the comprehensive list of UK businesses that is used by government for statistical purposes. It is fully compliant with the European Union Regulation on Harmonisation of Business Registers for Statistical Purposes. It fully complies with all European Union legislation relating to the structure and use of business registers, including:

- Regulation (EC) No 177/200814 of 20 February 2008 establishing a common framework for business registers for statistical purposes
- Council Regulation (EEC) No 696/9315 on statistical units for the observation and analysis of the production system in the Community; and 6
- Commission Regulation (EC) No 250/200916 as of 11 March 2009 implementing Regulation (EC) No 295 /2008 of the European Parliament and of the Council as regards the definitions of characteristics

The information used to create and maintain the IDBR is obtained from five main administrative sources. These are:

- i) HMRC VAT – traders registered for VAT purposes with HMRC
- ii) HMRC PAYE – employers operating a PAYE scheme, registered with HMRC
- iii) Companies House – incorporated businesses registered at Companies House
- iv) Department for Environment, Food and Rural Affairs (DEFRA) farms
- v) Department of Finance and Personnel, Northern Ireland (DFPNI)

As well as the five main sources listed above, a commercial data provider, Dun and Bradstreet, is used to supplement the IDBR with Enterprise Group information.

The IDBR is automatically updated by the data received from the following sources, with output files produced for areas where further clerical quality checking is required:

Daily updates

- VAT Traders File
- Companies House (Births and Deaths)

Weekly updates

- VAT 51s (Paper)

Companies House

- Fortnightly updates
- VAT Group Traders File

Monthly updates

- VAT turnover update

Quarterly Updates

- VAT turnover update
- PAYE update
- DEFRA update

Bi-Annual updates

- Vision VAT
- Redundant traders

Annual updates

- Intra/Extra Community Data update
- PAYE descriptions update
- Dun & Bradstreet

On a quarterly basis contact is made with HMRC PAYE to discuss the receipt and upload of the PAYE update. Twice a year a focus group meeting is held with Companies House. A minimum of two meetings per year is held with Dun & Bradstreet to discuss the dummy and live extracts. Additional data quality meetings take place if required. There are service level agreements and memorandum of understanding in place which are reviewed on a regular basis.

Imputation is used in cases where there is only a single source of admin data available for a business. This is either a VAT source or a PAYE source. Where a business is registered for PAYE only (that is, no VAT) the missing turnover variable is calculated using a Turnover per Head (TPH) ratio. Where a business is registered for VAT only then the employment is calculated using TPH. The TPH process is run once a year as part of the annual turnover update cycle in November.

Revisions are not applied. The Business Register is a live system and only represents the current picture.

Processing

This information is received in varying periodicities from daily through to annual updates, and is subjected to rigorous testing and quality control checks before it is uploaded onto the IDBR. Checks include, matching HMRC VAT and PAYE information, checking that business locations and structures match PAYE and VAT information, employment data are correct, businesses are active, allocating businesses to correct standard industrial classifications, etc. These tasks are carried out via automatic system checks, with changes and errors reported out for manual investigation and checking before correction and subsequent uploading.

There are also updates from ONS run surveys, such as the Business Register and Employment Survey, which provides classification, local unit and employment details.

A monthly quality report is produced for internal and external customers as part of the quality review process. The report provides details of quality issues identified during the previous month that SLA customers need to consider when using IDBR data. The data contains count, employment and turnover of all units on the IDBR and shows a comparison of this data on a month on month basis. The tables highlight any differences, which are investigated and commented on. Data is split by, region; SIC division, legal status and local unit count.

The quality statement of the IDBR, which can be used for sampling purposes, informs users of any system changes over the reporting period, and the impact this has had on the data. It is a key document to assess quality.

A dedicated team called the Business Profiling Team (BPT) is responsible for maintaining and updating the structures for the largest and most complex groups on the IDBR. BPT quality assures both the group structures and data (Employment, Turnover and Classification) for approximately 170 of the largest domestic and multinational enterprises (MNEs) every year. This profiling activity involves directly speaking to respondents of these groups either by telephone, e-mail or through face to face meetings to ensure that the legal, (administrative data), operational and statistical structure is accurately held on the IDBR. This ensures that high quality and timely statistical data is collated from these businesses via ONS's economic surveys.

The majority of the output files received on Admin Inputs are via WS_FTP Pro. The output files are collected daily and transferred onto Excel spreadsheets. To avoid loss of data, the files in WS_FTP Pro are not deleted until the excel spreadsheet has been created and quality checked. Other output files are found in the database which the team use to process the work. Banks and Building information is taken from the Bank of England webpage and is public knowledge. Academies data are collected from excel spreadsheets within GOV.UK/Publications /Academies. This information is available to the public. To help process the work, sites such as Companies House are used for confirmation; for example, company numbers, if companies are liquidated or dissolved. This information is updated daily and is available to the public.

The survey inputs team within the IDBR receives information provided by respondents via surveys, through information gathered from respondents by the Business Data Division, and from dead letters returned by Royal Mail. This information is taken and names, addresses, contacts, classifications and business structures are manually updated. The team also looks at gains and losses for surveys, which is a manual quality check on changes in employment and classifications that have happened on the IDBR since the last selection of the survey in question. There are also a number of other processes that quality assure data from survey sources, Companies House and HMRC, when those sources have impacted the structures and classifications of businesses on the Register.

In line with Continuous Improvement BRO is in the processes of reviewing its processes across all areas to ensure they are adding value to the quality of the IDBR. This is carried out annually.

Practice area 2: Communication with data supplier partners

The IDBR provides the main sampling frame for surveys of businesses carried out by ONS and other government departments. It is also a key data source for analyses of business activities. IDBR publications are:

- the annual publication "UK Business: Activity, Size and Location" (formerly known as PA1003 – Size Analysis of United Kingdom Businesses) provides a size analysis of UK businesses
- the annual publication "Business Demography" provides analysis on business birth, death and survival rates

Some customers have direct access to the IDBR, some will use published data, and some require bespoke analysis. These include:

- Welsh government
- Scottish executive government
- Business Energy and Industrial Strategy
- Department for Transport
- Department of Environment, Food and Rural Affairs
- Eurostat
- Department for Work and Pensions
- Health and Safety Executive
- Her Majesty's Revenue and Customs
- Environment Agency
- Scottish Environment Protection Agency
- Intellectual Property Office
- Department of Health

Practice area 3: Quality assurance principles, standards and checks by data suppliers

All data received is uploaded onto the IDBR and part of this process will involve going through a number of systems to ensure the quality of the information held and the company linkage is correct. Output files are produced where clerical investigation is required. All the systems have quality assurance and validation checks built in. These are different on each system.

Some are straightforward; for example, checking that the Daily VAT file and Daily CH file have the correct batch number (these have to be run in the correct order). Some are more complex; for example, for VAT numbers birthed in Aberdeen. Where possible all data received is validated; for example, the Standard Industrial Classification 2007 (SIC2007) code must be valid, or it will be amended to a default and reported.

Alongside the automated quality checks there is also a team of 17 administrators who, on a daily basis, quality assure the output files received from the IDBR with regards to VAT, PAYE, Companies House and the company matching process. The main function of the IDBR's teams is to carry out quality checking and updating of the IDBR. On a regular basis the managers carry out quality spot checking of the work carried out to ensure it is accurate. The team review around 141 different output files covering all aspects of the data received, ensuring they are accurately updated on the IDBR. On an annual basis the manager of the Enterprise Group team will quality check the test data received from Dun & Bradstreet to ensure it is fit for purpose prior to the upload of the actual data in January. The live extract is then clerically processed by the team.

A monthly quality report is produced for internal and external customers as part of the quality review process. The report provides details of quality issues identified during the previous month that SLA customers need to consider when using IDBR data. The data contains count, employment and turnover of all units on the IDBR and shows a comparison of this data on a month on month basis. The tables highlight any differences, which are investigated and commented on. Data is split by, region; SIC division, legal status and local unit count.

The quality statement of the IDBR, which can be used for sampling purposes, informs users of any system changes over the reporting period, and the impact this has had on the data. It is a key document to assess quality.

A dedicated team called the Business Profiling Team (BPT) is responsible for maintaining and updating the structures for the largest and most complex groups on the IDBR. BPT quality assures both the group structures and data (Employment, Turnover and Classification) for approximately 170 of the largest domestic and multinational enterprises (MNEs) every year. This profiling activity involves directly speaking to respondents of these groups either by telephone, e-mail or through face to face meetings to ensure that the legal, (administrative data), operational and statistical structure is accurately held on the IDBR. This ensures that high quality and timely statistical data is collated from these businesses via ONS's economic surveys.

Practice area 4: Producers' quality assurance investigations and documentation

For a general outline of QA procedures by us, which applies to IDBR, please refer to Annex B.

19. International Passenger Survey (IPS)

Practice area 1: Operational context and admin data collection

The International Passenger Survey (IPS) covers most large ports in the UK with shifts running at airports and St Pancras (for the Eurostar) all times of day and week. Boats are sampled at sea ports allowing for all times when they run. Four administrative data sources are used to weight up survey data to reflect the population: passenger numbers for flights are taken from Civil Aviation Authority (CAA) data, passenger numbers for sea travel from Department for Transport, and passenger numbers for the channel tunnel are from Eurostar and Eurotunnel in the provisional estimates. Administrative data can be delivered by individual airports for the monthly publication, but a complete data set is provided by the CAA for quarterly and annual publications; similarly, DfT provide final passenger numbers for the non-air routes.

Survey response was 77% in 2016. Most of the non-response is due to "clicks". This is where there is no interviewer available to administer the survey at busy times. These clicks are assumed to be completely random and similar to responding passengers. The survey interviews overseas residents who do not always speak English very well. We do have some language questionnaires to try and alleviate this. There are no coverage issues.

There is item non-response for some variables where respondents do not know the answers. Item non response is imputed using an iterative near neighbour method. Monthly outputs use some data from the year previous and calculate a factor to uplift traffic totals, although these are the extreme residual airports within the UK.

Administrative data is processed through excel workbooks for processing into the IPS weighting system.

Most workbooks have macros assigned to them to pick up the administrative data sources and add them to a time series workbook which then produces a graphical check for any large step changes in the data. If any large changes are identified then this is queried this with the suppliers. For those workbooks that do not operate on macros, data are manually copied and pasted, and formulae are copied down. The IPS team are working towards replacing manual steps with macros. The risk is minimal as all these workbooks have their own checks sheet with them checking totals from start to finish, entry period dates etc. Finally another final graphical check is in place to identify and large step changes after the processing of the admin data.

Practice area 2: Communication with data supplier partners

The following are users of IPS statistics:

- Bank of England
- Home Office
- DfT
- CAA
- Visit Britain
- National Accounts (Household Expenditure, Trade in Services)
- Migration Statistics
- HMRC
- numerous academics and travel consultants

Practice area 3: Quality assurance principles, standards and checks by data suppliers

All the data goes through vigorous checking. The administrative data are checked as they are received by adding the data to a time series, then checking for any step changes. Anything that seems to be unusual is queried.

Survey data checks are performed first by the coding and editing team, and secondly by the IPS research team to identify any further errors or edit queries. A number of frequency checks take place before the data is processed through the IPS weighting system.

Post-processing checks are carried out that look at any large weights, negative weights and missing weights. Weighted totals are checked against the input data (Admin data passenger totals) for air, sea and Eurostar trains. Finally a comprehensive breakdown of our publication reference tables is produced for quality assurance purposes, and a meeting is held for every monthly and quarterly publication to discuss the differences in trends. The publication is then signed off.

Revisions result from more accurate passenger figures being made available. Overseas travel and tourism monthly estimates are revised during the processing of the quarterly dataset and again during the processing of the annual dataset.

Practice area 4: Producers' quality assurance investigations and documentation

For a general outline of QA procedures by us, which applies to IPS, please refer to Annex B.

20. Kantar

Practice area 1: Operational context and admin data collection

Kantar collect prices data through a consumer panel of 15,000 individuals. These are in the age range 13 to 59, and are stratified into age, gender and region.

The consumer panel excludes Northern Ireland.

There is no contract or SLA in place for Kantar data. We purchase the data annually in a one-off payment.

Practice area 2: Communication with data supplier partners

There is a dedicated contact for any issues. When contacted, the Kantar representative agreed to a short telephone meeting to discuss their QA procedures. This was very productive and the representative provided thorough answers to the questions provided.

Practice area 3: Quality assurance principles, standards and checks by data suppliers

The classifications for stratification are matched with the stratifications used in our Census publication.

Individuals are given log on details to a system where they record information about their entertainment purchases. This data is collected every 4 weeks.

The weights data comes from expenditure recorded by EPOS, who provide the electronic point of sale technology to retailers. This data is collected by third party companies and purchased by Kantar.

Around 80% to 90% of retailers are picked-up by this data collection. A notable exception is Toys-r-us, who do not provide data. Therefore if there is a product that is exclusively stocked by this retailer, then it will not be represented.

This data is monitored over time. Once the data has been collected, there are several Quality Assurance Systems and processes in place.

Practice area 4: Producers' quality assurance investigations and documentation

For a general outline of QA procedures by us, which applies to Kantar, please refer to Annex B.

21. Moneyfacts

Practice area 1: Operational context and admin data collection

Moneyfacts produces a monthly magazine which lists price comparisons for several products, including mortgage arrangement fees.

Prices are collected from the "residential mortgages" section of the Moneyfacts publication. The subscription is delivered directly to Prices division around the beginning of each month. This is in the form of a hard copy and a PDF File.

Practice area 2: Communication with data supplier partners

The delivery of the Moneyfacts magazine is a subscription service, and there is no dedicated point of contact. If Prices division require clarification on figures, they will visit the Moneyfacts website, the address for which is provided in the magazine, and locate a suitable point of contact.

Practice area 3: Quality assurance principles, standards and checks by data suppliers

Moneyfacts has a research team which monitors mortgage products available in the UK. The mortgage data in the magazine is partitioned into the different companies, and each of these has a telephone number and website address.

The research team collects the data from the website

Information on quality assurance data procedures and quality assurance was not as readily available online as other administrative data sources. They do state that they aim for at least 95% of providers in their coverage, and that they are regulated by the Financial Conduct Authority.

There is some information that is provided by third parties. It is stated on the Moneyfacts website that these companies adhere to a code of conduct.

Practice area 4: Producers' quality assurance investigations and documentation

The following description of QA Procedures is taken from the Prices STaG documentation in our Prices division.

Processing redemption fees

Look for "Standard Redemption Conditions" for each sampled institution. This fee can be named differently depending on the bank. They can also be called discharge fees, booking fees, deeds fees, sealing fees, rdm admin fees; or a combination of these. As well as a fee, some banks have other redemption conditions relating to interest charges. We are only interested in the fees (or the fee combinations). Add the fees together if necessary, and enter the amount in the relevant cell in the "Redemption fees" worksheet. No extra work is required as the index calculates automatically.

Processing mortgage arrangement fees

If possible, collect the same mortgage as was collected in the previous month. Enter the mortgage description in the relevant columns of the "Mortgage Arr Fees" workbook, using information in Moneyfacts magazine (the description can be copied over from the previous month and modified if necessary). Enter the price of the mortgage. If the mortgage is the same, ignore the "N/C" (not comparable/comparable) column.

Sometimes the mortgage priced in the previous month is no longer for sale in the current month. If this happens follow the same procedures as in the paragraph above, but use the "N/C" column to indicate that the mortgage is not comparable by entering "N". The chosen mortgage should try to match the base product description as much as possible, though this will not always be possible.

How to determine whether or not a mortgage is not comparable

When assessing a mortgage product in the current month to determine whether or not it's comparable, the price analyst should examine it relative to the previous month's mortgage product. The mortgage market can be fast moving and the composition of mortgage products can also rapidly change.

The mortgage attributes which must be kept constant are: Buyer status (Mover, First Time Buyer or FTB, Remortgagor) interest rate type (fixed, tracker, variable, capped) term (length of time for which initial interest rate applies).

The following attributes should also be kept constant if possible, but some flexibility is allowed should a product change: whether or not there are early repayment charges the loan to value ratio (not always possible, but choose the next best 1 – some flexibility) it is preferable, also, to keep the information under "incentives/notes" in the publication, as constant as possible (though this is not a strict condition).

Interest rate information can be a useful guide for locating the comparable previous month's mortgage in the current month (some lenders offer many products and sometimes takes a long time to wade through), however they should be used with caution because rates frequently rise and fall. It is more important to look at the attributes rather than interest rates when selecting the mortgage product.

When the characteristics of a mortgage remain constant from 1 month to another but the interest rate changes, there is no truly objective way to judge whether or not the mortgage product is the same. To overcome this, a tolerance range was established. If the interest rate of a mortgage has increased or decreased +/- 0.5% compared with the previous month (not the base month), it can be considered to be a different mortgage (even if the characteristics remain the same or similar). This rarely occurs but is a good guideline should it happen. The price analyst should also examine BoE interest rate decisions as this will assist with deciding the degree of comparability between mortgage products.

For tracker mortgages (mortgages which track the BoE base rate plus a set percentage), the best way to locate a comparable mortgage in the current month is to find a product with the same set percentage rate (or as similar as possible). It may be that the base rate has increased or decreased but the mortgage should be considered comparable if the set percentage rate is the same as in the previous month and if the base rate is within the +/- 0.5% tolerance range.

Details of any unusual price movements, reasons for changes, or other points of interest, should be included in the "Data Notes" text box of the "Mortgage Fees Index" worksheet for future reference.

The base prices are the same prices that are requested for the January index plus the prices for any new items that are introduced for the new year. The price of each item is input into the base price column of the spreadsheet by the prices analyst making sure to match the prices with the row titles.

The prices analyst notifies the team coordinator that the spreadsheet is ready for checking and gives them the updated spreadsheet along with copies of the data to be checked. The team coordinator also updates the “Time Series Data” worksheet which contains the data used to generate the Time Series graph on the “Mortgage Fees Index” worksheet.

Once the spreadsheet has been checked by the team coordinator, it is passed on to the designated spreadsheet sign-off checker, along with the base prices, documentation and any new weights data.

The fees types priced should be reviewed regularly to ensure the index estimates remain correct. New fee types often emerge and the extent to which these fees affect mortgagees needs to be established. Likewise, other fees may drop out.

22. Websites

Background to data

The Consumer Prices Index including owner occupiers’ housing costs (CPIH) shopping basket consists of over 700 items. The majority of prices for these items are collected by TNS, a company that employs field collectors to visit stores around the country and check the prices of items within shops every month; these are then compared with the base January price when calculating the index.

Not all prices can be collected this way. There are some elements of the basket which are more akin to services, which cannot be stocked on a shelf. Examples of this include water and wastage services and child care. For these, prices are centrally collected by either viewing prices on a website, or directly contacting a service provider each month to determine if their price has changed.

Under the established definition of admin data, this would be classified as an administrative data source, as it is information originally collected for non-statistical purposes, which is then acquired and used for statistical purposes.

Information obtained from websites are under the technical definition, administrative data, however here are two main aspects to consider when determining whether to apply the full QAAD assessment on these sources:

- the weights are small, and therefore provide a minimal effect on the CPIH index
- the resources involved in conducting a full QAAD assessment on every website or direct contact, and ensuring these standards are kept in the future, far outweigh what contribution the sources have, and would be beyond the capabilities of prices division

Nevertheless, the sources are used in the production of an important economic index, and therefore some level of assessment is required. The QAAD assessment will be conducted on the general acquisition and use of website information, rather than individually.

Practice area 1: Operational context and admin data collection

At the beginning of the year, market share data of shops is obtained, usually through Mintel. This information is used to select which shop websites should be used to obtain price data.

Practice area 2: Communication with data supplier partners

The prices are collected from company websites. There is therefore no contract or point of contact established. If an item becomes unavailable, for instance due to the website not being accessible, then the item price will be imputed.

If an entire website were to become unavailable, then the item would be listed as out of stock.

If this issue persisted for several months, it would be treated as if the store had closed, and the next shop on the Mintel Market Share list would be used instead.

Practice area 3: Quality assurance principles, standards and checks by data suppliers

Prices are collected individually from websites. There is no available Quality Assurance on what online stores have in place to prevent pricing errors.

Practice area 4: Producers' quality assurance investigations and documentation

When the prices have been collected, a PDF of the price page is taken. Prices staff use this to check for any obviously incorrect data using their own expert knowledge and judgement.

The Pretium system, which is used to process the index, has a built in check error fail, which will flag up any values that fall outside a threshold.

If there is no way of knowing the weights of an item that has been collected, an estimate is taken.

It is acknowledged that due to resource limitations, the samples may be small on occasion.

The price is then inputted in to the main CPIH calculation.

Prices Production areas are externally accredited under the quality standard ISO9001, which promotes the adoption of a process approach, which will enable understanding and consistency in meeting requirements, considering processes in terms of added value, effective process performance and improvements to processes based on evidence and information. These standards are adhered to when collecting from websites.

Notes for: Annex A: Assessment of data sources

1. Includes intermediate tenures and other tenures not socially or privately rented
2. Not yet published

8. Annex B: ONS data checking and validation

Scrutiny is the name for a series of computerised checks carried out after price data is taken on to the database. The checks are designed to identify prices with large price changes and to manually consider their validity or identify those that are well outside the usual range of prices for a product. Credibility is a more refined computer test which uses a "Tukey algorithm" designed to identify outlying prices within each item. The terms scrutiny and credibility are also used to refer to the stage of work in the monthly cycle where the staff examine and deal with prices that have failed these tests.

Here is a link to the timetable for the various scrutiny and other validation actions that take place during the month and which are described below.

Responsibility for data checking and validation is divided between the Assistant Prices Managers according to the type of item. Both Assistant Prices Managers and Price Analysts are responsible for assessing price quotes flagged as part of scrutiny or validity.

EO Assistant Prices Manager responsibilities by group

Consumables and Transport

- 21 - Food
- 22 - Catering
- 31 - Alcohol
- 32 - Tobacco
- 62 - Fares and Other Travel Costs

Services

- 41 - Housing
- 42 - Fuel and light
- 44 - Household services
- 5203 - Personal services
- 64 - Leisure services

Durable Goods and Motoring Expenditure

- 43 - Household goods
- 51 - Clothing and footwear

1.1 Scrutiny checks

Three checks are carried out on the data by the computer as part of scrutiny process:

- min – max check
- monthly percentage change check
- invalid use of indicator codes

1.2 Credibility checks

The Tukey algorithm looks through the data and identifies outliers. The “Quote status” indicator shows which prices have failed the credibility check.

1.3 Dealing with invalid quotes

Price analysts examine every price failing scrutiny or credibility and decide what action should be taken. To decide which action should be taken on an individual price the price analyst will usually need to look at the Indicator Code, the history of the quote, the quote description, the Index, any messages left by the collector and previous queries raised with TNS in accordance with the processing checklist. The computer manager runs an Open Road process called “re-accepting fruit and veg quotes” to automatically accept items for some notoriously volatile prices experienced for fruits and vegetables.

1.4 Raising queries with TNS

If there is not sufficient information available to determine the appropriate action for a quote, a query to TNS is raised. This is a request asking for more information about the quote. Queries to TNS are raised on the system as part of the validation process, all queries should be extracted by the Processor (or nominated representative) and sent to TNS on a regular basis. TNS should respond to all queries within 3 days of receipt.

1.5 Other checks

By Briefing day all data checking and validation should be completed and all amendments made so that the first briefing prints showing the item indices can be produced. The Prices managers will then carry out index error (Howler) hunts looking at item indices for any obvious anomalies.

The following instructions cover the range of checks and validations that are carried out as part of the monthly cycle. A timetable is used to track these operations. The timetable containing key dates is circulated each month for reference by the processor (usually Computer Manager).

2.1 Scrutiny checks

The computer carries out three checks as part of scrutiny process. Once the scrutiny checks have been carried out a “Quote Status” code is added for each price. This code indicates whether the price has passed the scrutiny tests and is valid, or if it is invalid the reason that it failed the scrutiny tests. A list of the status codes is given below. A decision diagram shows the process in Annex A. When the data checks have been completed, scrutiny reports are printed which list each quote which fails the tests.

Table 7: Scrutiny checks

Status	Description
0	Price is outside the min-max range
1	Zero price or failed credibility check
2	Rejected by user
3	Validated

- 4 Accepted by user
 - 5 Price change failed % test
 - 6 Not currently used
 - 7 Unknown Indicator code used
 - 8 Indicator Code = Q/C/N/W but no message exists
 - 9 Price is 0 but Ind. is NOT T or M
 - 10 Indicator Code is T or M but price is NOT 0
 - 11 Quote is valid but Ind. is Q/W
-
-

Source: Office for National Statistics

2.1.1 Min-Max check

This check looks to see whether the price lies within the expected price range for the item. Any prices which lie outside the range will fail and be assigned a status code of 0. This check is carried out automatically by the computer system. Those prices which lie outside the range will need to be validated via the scrutiny reports which are explained in section 4.1.4. At the beginning of the year the minimum and maximum price for each item is set by the Assistant prices manager based on past experience of the likely range of prices for similar items and anecdotal evidence. A minimum and maximum price for the following months is derived from the minimum and maximum valid prices from the previous month, that is, the band is automatically widened as prices that were previously outside of the banding get accepted. This test is not carried out on fresh fruit and vegetables due to the large monthly price movements that these items experience.

2.1.2 Monthly percentage change check

This check is carried out automatically by the computer system and calculates the price change between two successive months' prices. The price will fail the check if the change is larger than the agreed maximum percentage change range for that item and it will be assigned a status code of 5 and will then need to be validated via the scrutiny reports.

The agreed percentages are:

Clothing and footwear: 40%

Food: 35%

Home-killed lamb: 50% Fresh fruit and vegetables: 0% (Prices for these items cannot fail percentage test)

All other items: 33%

2.1.3 Invalid use of indicator codes

Some prices are failed because of indicator code problems. Single letter indicator codes are used by the price collectors to provide information on price changes that have occurred and to give us additional information. A list of indicator codes is given in the following table:

Table 8: Indicator codes

Indicator Code	Definition
S	Sale price or special offer
R	Recovery of price at the end of sale or special offer
T	Temporarily out of stock
M	Missing - not sold at shop
C	Comparable - change in product being priced, new product is similar to the previous product
N	Non-comparable - change in product being priced, new product is not comparable to previous product
Q	Additional comments have been supplied
W	Size or weight change notified following a request for information in the collectors newsletter
X	Comparable Item On Sale
Z	Non-Comparable Item On Sale

Source: Office for National Statistics

There are five possible reasons for failure;

I. Unknown indicator code: A character has been keyed in the indicator code field that is not one of the 10 listed or is in lower case. This will be assigned a status code of 7.

II. Indicator is Q/C/N/W but no message exists: Some indicator codes require the collector to enter a message, if no message is added the quote will fail, and will be assigned a status code of 8.

III. Price is Zero but indicator code is not "T" or "M": If an item is not stocked in a shop or is temporarily out of stock the collector should enter a "T" or "M" in the indicator code field. If the collector has not used an indicator code the quote is rejected, and will be assigned a status code of 9.

IV. Indicator code is "T" or "M" but price is not Zero: This is basically the reverse of the above. A price has been entered and an indicator code "T" or "M" has been used. This should not happen as both mean the item is unavailable. This will be assigned a status code of 10.

V. Quote is valid but indicator is Q or W: Occasionally, collectors will be asked to check certain items to see if there has been a change to the size or weight, or they may wish to give some further details about the price and will use either Q or W. However, the price may still have passed the other two data checks. This will be assigned a status code of 11.

Any quotes which pass all three scrutiny check will be assigned a status code of 3. If any quotes are zero, that is, no price has been supplied and they have a T or M indicator code, they are assigned a status code of 1.

2.1.4 Scrutiny report

A Scrutiny report is produced showing all the quotes that have failed the computerised checks. It contains the following components:

PROCESSING PERIOD – Month and year we are currently producing figures for.

PRODUCED ON – Date and time the report was produced.

SECTION – Name and identification number of the group of items you are currently working with, for example, 2109 POULTRY.

ITEM – Each item is given a six-figure identification number called an “ITEM ID”. You can use the Item ID to find out which “GROUP” and “SECTION” the item is included in. The Item ID breaks down as follows:

Table 9: Coding example

ITEM	210905	FRESH/CHILLED CHICKEN 1.4 KG
SECTION	2109	POULTRY
GROUP	21	FOOD

Source: Office for National Statistics

LOCATION – Prices are collected around 150 locations across the UK. Each of these locations is given a four or five-digit code called a “LOCATION CODE”.

SHOP – Within each location every shop is given a “SHOP CODE”.

SHOP NAME – Name of shop where the price was collected.

PRICE – Price for current month input onto collection device by the price collector.

IND. – Indicator code input onto collection device by price collector.

STATUS – Status code given to each item following scrutiny checks carried out during take-on programme.

STATUS DESCRIPTION – Description of rejection reason.

2.2 Credibility checks

Credibility is a more refined computer test which uses the Tukey algorithm to identify outlier prices within each item. The “Quote status” indicator shows which prices have failed the credibility check.

Once credibility has been run, the invalid prices from central shops are checked by the Prices analysts.

The Tukey algorithm

The Tukey algorithm identifies and invalidates price movements which differ significantly from the norm for an item. For seasonal items where price movements are erratic, the algorithm looks at price levels rather than price changes. The algorithm operates as follows:

- the ratio of current price to previous price (price relative) is calculated for each price
- for each item the set of all such ratios is sorted into ascending order and ratios of 1 (unchanged prices) are excluded
- the top and bottom 5% of the list are removed
- the midmean is the mean of what is left, as per Technical Manual
- the upper and lower semi-midmeans are the midmeans of all observations above or below the median
- the upper Tukey limit is the midmean plus 2.5 times the difference between the midmean and the upper semi-midmean
- if the upper limit is negative it is set to zero
- price relatives outside the Tukey limits are flagged for manual scrutiny

2.2.1 Big Shop – algorithm failures

Special reports are produced by the computer manager for the Big Shops both centrally and regionally collected as they carry a high weight and changes in these items have a greater impact on the overall index. Printouts for the Big Shops are produced and circulated after Credibility has been run.

2.2.2. Briefing reports

The Briefing report is the most important print for the Assistant prices manager. It shows the month's index for each item along with the index for last month and the same two months of the previous year for comparison, the monthly changes for the current year and the previous year and the contribution to the 12 month change for each item. Briefing reports are created by downloading item indices information from an Excel spreadsheet, the "Briefing Data Master" to produce a print. The briefing report also contains details and an explanation of significant changes to the index in the current year and the previous year.

2.2.3 Shop reports

Shop reports are run by the AO (Price Analyst) to aid explanation of price movement and where they have occurred. By looking at the shop reports you can identify where particular shops have price increases or sales etc. or where price changes are restricted to certain brands. The shop report does not show the item description but if you see the same price change for the same item in a large number of outlets you can then look up some of the quotes on the Retail Price Index (RPI) database to identify the brand. To run a shop report; from top menu in Open Road (rpi live), select analysis reports; shop report; double click item in list, checking item appears on the right-hand side of the screen. Once all required items have been selected click OK. All reports will print automatically, collected as a private job via a local printer, access code 0580.

2.3 Other checks

2.3.1 Data checks

To ensure that prices are not omitted from the index indefinitely, the computer system imputes base prices where the price has been missing or invalid for three consecutive months. Each month a report of all such quotes is produced to enable an Assistant prices manager to consider the validation of the quotes to see if the imputation can be avoided therefore maintaining the real price chains where possible. This is usually carried out on Wednesday during Briefing Week. The computer manager also runs an sql to pick up all invalid quotes therefore ensuring that nothing has been missed during scrutiny checks.

2.3.2 Index error (Howler) hunt

This check looks at any valid price quotes that show a large price relative (compared to the base period). It acts as a final check on the acceptance of high or low level matched pair (quote and base indices) and identifies those either lower than 60 or higher than 180. The Prices manager can accept or reject the quote from the final calculation as required. The process is run by the Prices manager within an excel spreadsheet once credibility runs have been completed, this is usually just prior to the Briefing meeting.

2.3.3 New/non-comparable quote report

This check looks at whether new items are comparable rather than non-comparable. A list of all products with an "N" indicator code is produced and these are reviewed to see if some should be reclassified as a comparable.

2.3.4 EO check

Once the briefing report and actions resulting from the briefing are complete and all indices for the current month have been fed onto the mainframe and shortly before Closedown, the EO Check Report is produced by the processor. The processor passes the relevant sections to the Assistant prices managers who then check that the data on the reports matches the data on the monthly spreadsheets or Pretium system.

9. Annex C: Summary of monthly validation

	Monday	Tuesday	Wednesday	Thursday	Friday
Week 1/5	PRE-COLLECTION DAY SPREADSHEET COLLECTION	COLLECTION DAY SPREADSHEET COLLECTION	POST COLLECTION DAY SPREADSHEET COLLECTION		

				1ST TNS DATA HOPEFULLY INPUT SPREADSHEET DATA CENTRAL SHOPS DATA COLLECTION	1ST TNS DATA TIMETABLED CONTINUE WITH SPREADSHEETS CENTRAL SHOPS DATA COLLECTION
Week 2	FINALISE SPREADSHEETS INPUT CENTRAL SHOPS DATA	2ND TNS DATA HOPEFULLY SCRUTINY SORTED START PASSING SPREADSHEETS TO TEAM CO- ORDINATOR	CENTRAL SHOP DEADLINE SCRUTINY QUOTES AVAILABLE FOR ACTION CREDIBILITY TO BE RUN TONIGHT	QUERIED QUOTES BIG SHOP FAILURES	INVALID QUOTES FOR ACTION ALL SPREADSHEETS TO TEAM CO- ORDINATOR BIG SHOP FAILURES
Week 3	FINISH ACTIONING INVALID QUOTES ACTION QUERY RESPONSES START PASSING SPREADSHEETS TO SEO/HOB	SCRUTINY DEADLINE	WORK ON BRIEFING REPORTS	OUTSTANDING SPREADSHEETS TO SEO/HOB WORK ON BRIEFING REPORTS PRODUCE HOWLER HUNT	BRIEFING DAY HOWLERS
Week 4	BRIEFING DAY ACTIONS FINISH HOWLERS	CLOSE DOWN EO CHECKS	OUTLOOK TO BANK OF ENGLAND IN MONTHS WHEN THE MPC IS SITTING IN THE WEEK PRIOR TO PUBLICATION	RPIJ/RPIY/TPI /CPIY/CPIC /CPIH FINALISED STATISTICAL BULLETIN PREPARED	ADDITIONAL BRIEFING FINALISED FINALISE THEME PAGE SUMMARY STATISTICAL BULLETIN TEXT TO TRIDIAN
Week 5/1	STATISTICAL BULLETIN, THEME PAGE AND ADDITIONAL BRIEFING TO PRESS OFFICE HMT MEETING	PUBLICATION DAY AGENCY AND PRESS BRIEFINGS	QUALITY DAY	MONTHLY REVIEW MEETING	
	<<<<	<<<<		>>>>	>>>>

10. Annex D: Quality assurance processes and checks – UK Consumer Research

Within Mintel, the UK Consumer Research and Data Analytics team (CRDA) is responsible for ensuring the quality assurance of consumer data across UK Mintel reports and other published content.

Mintel are full members of the UK Market Research Society (MRS) and adhere to MRS guidelines and codes of conduct in regards to all aspects of the quantitative and qualitative data collection.

This document details our quality processes and checks for the following:

- online quantitative research – Lightspeed GMI
- face-to-face quantitative research – Ipsos MORI
- online and face-to-face quantitative data checks – Mintel CRDA team
- online qualitative research – FocusVision Revelation
- Mintel forecast
- data collection auditing

Online quantitative research – Lightspeed GMI

The majority of our online quantitative consumer research is conducted using panel from Lightspeed GMI. The process below details the quality assurance checks conducted at each stage from questionnaire design to reporting.

Questionnaire design checks

The CRDA team contains questionnaire design experts who work with industry analysts to produce the best content possible in surveys designed for optimum engagement and quality collection of data. Surveys are quality checked by a manager within the report industry team as well as an independent checker from within CRDA.

Internal scripting

Our in-house team script questionnaires on FocusVision's Decipher platform. A leading scripting tool used by major agencies and fieldwork suppliers.

Mintel has its own internal survey scripting resource which sits within our CRDA team. Rather than commission panel providers to script our surveys we have chosen to retain control over this aspect of our research process. This allows the CRDA team to have a direct influence on how surveys look and feel as well as being able to resolve any survey queries quickly and effectively.

Our survey scripters have their own internal quality checks in place, ensuring each project they work on is checked for errors by another team member before it is sent back to industry analysts and the rest of the CRDA team. Scripters also work together with our panel provider Lightspeed GMI to monitor the look and feel of our surveys to ensure we produce "best in class" scripts. This includes utilising aspects such as gamification, iconography and pictures to drive up survey engagement.

Test link checks

Once scripted, at least two members of the CRDA team, together with the industry analyst(s) who commissioned the questionnaires, test the survey link to ensure what is reflected on the final document is accurately shown on screen.

At this stage, we also test the time taken to complete. Our UK questionnaires are designed to be no longer than 15 minutes in length. The vast majority of our surveys fall below this benchmark (median time currently standing at 11 minutes as of March 2016). Maintaining survey lengths of below 15 minutes ensures we do not compromise the quality of our data from the effects of respondent fatigue.

Dummy data checks

Once the CRDA team and analysts have signed off a link we run a set of 500 dummy completes through the survey script. The CRDA team then checks the results to ensure routing and filtering runs correctly. This also allows us to check that inputs feed into the data map as required.

Soft launch and pilot checks

Once dummy checks have been completed the survey is piloted to 100 live respondents. Once 100 responses have been achieved we pause the wave to re-check routing and filtering as well as checking drop-outs by question. If a particular question(s) cause concern we will investigate and re-format if needed.

Fieldwork, data validation and processing checks

Before and during fieldwork Lightseed GMI perform checks on their panel which prevents fraudulent respondents joining and entering surveys whilst also removing over-reporters and eliminating duplicates. The following details the validations and checks which are performed:

1. Identity validation

Identity validation is performed at recruitment, as a respondent joins the panel. This is done through matching PII information (one or combination of name, physical address, email address) to a third-party database.

2. IP address validation

IP address validation is conducted at panel registration or before entering surveys from other sources. This is done through validating the country and region of origin of IP, detection of proxy servers and against a known list of fraudulent servers.

3. Honesty detection

Lightpeed GMI's "Honesty Detector" system is used prior to a respondent leaving the panel system to start an external survey. This system is used to analyse the respondent's responses to a series of low incidence activities and benchmark questions. This identifies unlikely combinations of responses to detect outliers in data and removes over-reporters. Respondents are certified as "Honest" every 60 days.

4. Unique survey responders

At the start of a survey Lightpeed GMI uses proprietary and industry standard digital fingerprinting tools to identify and eliminate duplicates from the study. This works in conjunction with IP address validation, where permitted. This identifies respondents who have already accessed a survey from any incoming source and prevents them from entering twice.

In addition to these checks Mintel also carries out checks for "speedsters" – respondents who complete the survey in a time deemed too quick. Respondents failing the speedster check are removed from the data set.

Throughout fieldwork checks are made daily by Mintel to monitor drop-out rates and participation. Furthermore, completion rates against our quotas are also constantly monitored and managed by Lightspeed GMI. At the end of fieldwork the number of completes is assessed against our quotas. Up to a 5% deviation can be applied to the number in each quota cell if required. Respondents are incentivized for taking the survey in the form of points which are allocated and managed by Lightspeed GMI.

Excel tables and an SPSS file are produced for each wave. These are formatted to a pre-agreed specification by the scripting team. The tables generated are quality checked by a member of the scripting team against a raw data file. The SPSS file is then checked against a code book and the tables. Please see the section "Online and face-to-face data checking – Mintel CRDA Team" for a detailed outline of the next stage in the process.

Face-to-face quantitative research– Ipsos MORI

Questionnaire design checks

The CRDA team contains questionnaire design experts who work with industry analysts to produce the best content possible in surveys designed for optimum engagement and quality collection of data for a face-to-face methodology. Surveys are quality checked by a manager within the report industry team as well as an independent checker from within CRDA.

Once approved by the CRDA team the questionnaire is sent to the Ipsos MORI Capibus team who perform a further quality check before the questionnaire is processed into their CAPI system.

Sampling

Ipsos MORI's Capibus uses a rigorous sampling method – a controlled form of random location sampling (known as "random locale"). Random locale is a dual-stage sample design, taking as its universe sample units, a bespoke amalgamation of Output Areas (OAs – the basic building block used for output from the Census) in Great Britain. Ipsos MORI uses a control method applied to field region and sub-region to ensure a good geographical spread is achieved.

Stage one – Selection of primary sampling units: The first stage is to define primary sampling units (PSUs). Output areas are grouped into sample units taking account of their ACORN characteristics. A total of 170 to 180 PSU's are randomly selected from our stratified groupings with probability of selection proportional to size.

Stage two – Selection of secondary sampling Units: At this stage, usually two adjacent output areas (OA), made

up of c.125 addresses each, are randomly selected from each primary sampling unit, this then becomes the secondary sampling unit. Interviewers are set quotas for sex, age, working status and tenure to ensure the sample is nationally representative – using the CACI ACORN geo-demographic system in the selection process. Using CACI ACORN allows Ipsos MORI to select OAs with differing profiles such that they can be sure they are interviewing a broad cross-section of the public. Likelihood of being at home and so available for interview is the only variable not controlled for. Fieldwork times and quotas are therefore set to control for this element – age and working status and gender – giving a near to random sample of individuals within a sample unit. Typically Ipsos MORI use 170 to 180 sampling units (sampling points) per survey. Precise sampling units of addresses combined with control of quotas affecting likelihood of being at home produces a sample profile that is similar to that achieved on The National Readership Survey (which uses random probability sampling) after four call-backs. Only a limited amount of corrective weighting is therefore needed to adjust the final results so that they are in line with the national demographic profile.

Interviewing

Interviewing occurs between 1pm to 9pm, with 50% conducted during the evening and at weekends and 50% conducted during week days. Ipsos MORI has around 1,500 interviewers in the UK and the Republic of Ireland. Their large field force means that they can have locally based interviewers who have a detailed knowledge of, and sensitivity to, the local area. Since they do not rely on sub-contracting their fieldwork, they can ensure that quality standards are observed consistently at every stage. Participants are not given an incentive for taking the survey.

Ipsos MORI's large field force enables a spread of interviewers to be used to minimise bias in the responses and also to minimise risk to the data and delivery of the project if any issues are raised with a particular interviewer's work.

Data validation

Validations are carried out using CATI (Computer aided telephone interviewing) within Ipsos MORI's telephone centres. A specially trained team of validators is used with 10% of all the validations monitored by a supervisor. Any interviews carried out in a language other than English, are validated by the CATI validation team in the same language.

Questionnaire data is received and loaded into the Field Management System (FMS). Sample is then transferred electronically into their telephone dialing system. Special validation scripts are used to ask questions to ensure that the interview was carried out professionally, in the proper manner and that key demographics are recorded accurately. They include several additional project specific questions, which can also check accuracy against the recorded data. Approximately 15 questions are asked or checked during each validation. Although the majority of validations are carried out using CATI, a proportion are carried out using a postal validation questionnaire, where a telephone number is not recorded or where an attempt to call by telephone results in a wrong or unobtainable number or repeated no answer. Occasionally a personal validation is carried out where a phone number is not given or when there is a concern as a result of a telephone validation. This involves a supervisor re-visiting respondents in person in order to validate the interview process.

Data checking – Ipsos MORI

Computer tables and an SPSS file are produced, formatted to Mintel's specification. The tables generated are quality checked by a member of Ipsos MORI's team against a raw data file.

The SPSS file is checked against a code book and the tables.

All information collected on Capibus is weighted to correct for any minor deficiencies or bias in the sample.

Capibus uses a "rim weighting" system which weights to the latest set of census data or mid-year estimates and NRS defined profiles for age, social grade, region and working status – within gender and additional profiles on tenure and ethnicity.

Rim weighting is used to provide the "best weighting", or least distorting, by using computing power to run a large number of solutions from which the best is chosen. Thus "Rim weighting" is superior to the more common system of "Cell weighting".

Online and face-to-face quantitative data checks – Mintel CRDA Team

All survey data (including any weighting) is checked against the top lines supplied by our in-house scripting team or external research partners. This is the first step in our process before any data analysis begins.

We use a variety of analysis techniques in SPSS as well as a custom version of FocusVisions' Decipher reporting

tool for simple crosstabs and significance testing. Typically members of the CRDA team will work to produce outputs in a standardised format so that should any mistakes arise, these can be easily spotted and rectified. Each piece of analysis is checked by someone in the CRDA team other than the person who worked on it initially. For more complicated analysis, a third checker within the CRDA team will also be involved. The CRDA checker will check all files related to the analysis including the SPSS syntax, data, initial analysis request form and the final deliverable charts and/or report section.

Report analysts also have limited access to a version of the data via the Decipher reporting tool where they can create simple crosstabs and custom groups using survey questions. All crosstabs produced by industry analyst teams are quality checked by a member of the CRDA team before being added to the report data book.

We have strict conditions over who has access to each set of survey data on the reporting tool. Regular checks are in place to ensure that only those report analysts working on a particular report will have limited access to the survey response data.

Once the report is ready for publication, all sections are quality checked in detail by a manager within report industry teams for example, finance, retail, food and drink, and so on). Any data queries are investigated and rectified before the report moves into the proofing stage. The proofing team are not involved in the report's content generation and hence, are in a unique position to sense check the report for logic, grammar and spelling issues.

Online qualitative research – FocusVision Revelation

FocusVision provides Mintel with qualitative bulletin board software "Revelation". This allows the creation of Internet-based, "virtual" venues where participants recruited from Mintel's online quantitative surveys gather and engage in interactive, text-based discussions led by Mintel moderators.

Discussion guide creation

The CRDA team contains qualitative discussion guide experts who work with industry analysts to produce the best content possible in order to reach research objectives. Qualitative discussion guides are quality checked by a manager within the report industry team as well as an independent checker from within CRDA.

Sample

Participants are recruited to online discussions from Mintel's online quantitative surveys (through Lightspeed GMI's panel). A question is included at the end of each of our quantitative studies asking for the participant's agreement to be re-contacted to take part in future Mintel online discussion groups within the following 3 to 4 weeks.

If the participant takes part in a follow up project they are incentivized in the form of points which are allocated and managed by Lightspeed GMI.

Online fieldwork and moderation

Discussion guides are uploaded to the Revelation portal, where they are again further checked for accuracy from a second CRDA team member. Once started, our discussions last for no longer than 5 days during which moderators are available to answer queries and ensure participants are adhering to rules surrounding interactions with others and their general use of language. Additionally, FocusVision provide a 24-hour help desk where any potential incidents or problems occurring outside of UK office hours can be investigated and resolved.

Transcript creation and checking

At the end of fieldwork the CRDA team downloads the discussion transcripts from the Revelation portal. These are checked for accuracy against the online discussion. Personal identifiable information is stripped out from the transcript files before sending to industry analysts for analysis.

Reporting

Within Mintel Reports Industry analysts can choose to use selected extracts from relevant qualitative discussions – this is shown as verbatim. To ensure Mintel conforms to MRS/ESOMAR guidelines we ensure that the participant's right to anonymity and confidentiality is respected and protected. Verbatim used in Reports can only be followed by a basic demographic profile (for example gender, broad age group, socioeconomic grade and so on).

Removal of discussions or personal data

The discussion and associated personal identifiable information associated with it are completely removed from the FocusVision Revelation platform and internal Mintel storage systems 3 months after the end of fieldwork.

Mintel forecast

For the Mintel forecast, the most appropriate statistical forecast is selected based on a market brief provided by our in house report analysts who specialise in a variety of markets. Like all other analyses, this is second checked by another member of the CRDA team and then signed off by our report analysts based on their knowledge of trends in that particular market.

Data collection auditing

Online – Lightspeed GMI

UK online data collection is audited every 3 months. During this audit we perform the following checks and procedures:

- monitoring drop-outs rates across our waves – online waves must not have a drop-out rate of more than 10%; those exceeding that figure will be assessed for data quality
- monitoring median completion times – online waves should fall below a median completion time of 15 minutes
- monitoring time in field – online waves should not exceed over 15 days in field
- cleaning of personal identifiable data – personal data collected through our online waves (that is, first, last names, email addresses and other sensitive information) is cleaned out of our system once 3 months has past from the end of fieldwork or data collection

All methodologies

Once a year we perform the following:

- methodology audit – we assess our methodology against recent analyst and client feedback as well as looking at research industry best practice and trends from the ONS, ESOMAR and other sources
- survey quota update or audit – we update our online survey quotas yearly using the most up-to-date data available from the Office of National Statistics for the distribution of gender, age, region and socioeconomic grade in Great Britain; our age quotas are weighted against internet penetration in each age group to be representative of the internet population in Great Britain
- supplier audit – we audit our online data collection suppliers yearly to assess that they still conform to industry standards, we also reassess their panel size and quality; for new suppliers we ensure they can fully answer and conform to “ESOMAR 28” - a standard set of questions research buyers can ask to determine whether a sample provider’s practices and samples are fit for purpose
- data protection or storage audit – our internal data security manager assess our polices and producers regarding data collection and storage of personal identifiable data; all quantitative and qualitative data is collected and stored on UK-based servers, satisfying European data protection law

11. Annex E: Flow diagrams of quality assurance processes

[Annex E: Flow diagrams of quality assurance processes](#)