

# Quality assurance of administrative data used in consumer price inflation statistics

Quality assurance (QA) assessment of all data sources used in the production of the Consumer Price Index including owner occupiers' housing costs (CPIH), to include the used cars data acquired from the "Autotrader Group".

Contact:  
Prices Division  
cpi@ons.gov.uk  
+44 1633 651976

Release date:  
26 March 2025

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](#)
- . [Annex A: Assessment of data sources](#)
  - . [Annex B: ONS data checking and validation](#)
  - . [Annex C: Quality assurance processes and checks – UK Consumer Research](#)
  - . [Annex D: 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, Welsh and Scottish governments and the Northern Ireland Housing Executive. Data from a range of sources 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

The details below summarise the quality assurance benchmarks that were set, and the assurance levels that we have assessed each source at during this assessment.

For used cars Autotrader, the:

- risk was low
- profile was high
- benchmark QA level was enhanced
- achieved assessment is still in progress

For RDG LENNON, the:

- risk was low
- profile was high
- benchmark QA level was enhanced
- achieved assessment is still in progress

For Valuation Office Agency rental price data, the:

- risk was medium
- profile was high
- benchmark QA level was comprehensive
- achieved assessment was comprehensive

For Welsh Government rental price data, the:

- risk was low
- profile was medium
- benchmark QA level was enhanced
- achieved assessment was comprehensive

For Scottish Government rental price data, the:

- risk was low
- profile was medium
- benchmark QA level was enhanced
- achieved assessment was comprehensive

For Northern Ireland Housing Executive rental price data, the:

- risk was low
- profile was medium
- benchmark QA level was enhanced
- achieved assessment was comprehensive

For Mintel, the:

- risk was medium
- profile was medium
- benchmark QA level was enhanced
- achieved assessment was comprehensive

For Glasses, the:

- risk was low to medium
- profile was low
- benchmark QA level was basic
- achieved assessment was basic (incomplete)

For Moneyfacts, the:

- risk was low
- profile was low
- benchmark QA level was basic
- achieved assessment was basic

For HESA, the:

- risk was low
- profile was high
- benchmark QA level was basic
- achieved assessment was basic

For Consumer Intelligence, the:

- risk was low
- profile was low
- benchmark QA level was basic
- achieved assessment was basic

For Kantar, the:

- risk was low
- profile was low
- benchmark QA level was basic
- achieved assessment was basic

For IDBR, the:

- risk was low
- profile was low
- benchmark QA level was basic
- achieved assessment was basic

For Website, the:

- risk was low
- profile was low
- benchmark QA level was basic
- achieved assessment was basic

For Direct Contact, the:

- risk was low
- profile was low
- benchmark QA level was basic
- achieved assessment was basic

For Brochures, the:

- risk was low
- profile was low
- benchmark QA level was basic
- achieved assessment was basic

For HHFCE, the:

- risk was high
- profile was high
- benchmark QA level was comprehensive
- achieved assessment was enhanced

For LCF, the:

- risk was low
- profile was high
- benchmark QA level was enhanced
- achieved assessment was comprehensive

For TNS, the:

- risk was medium
- profile was high
- benchmark QA level was comprehensive
- achieved assessment was comprehensive

For BEIS, the:

- risk was low
- profile was low
- benchmark QA level was basic
- achieved assessment was basic to enhanced

For IPS, the:

- risk was low
- profile was low
- benchmark QA level was basic
- achieved assessment was basic to enhanced

For Home and Communities agency, the:

- risk was low
- profile was low
- benchmark QA level was basic
- achieved assessment is still in progress

For the Department of Transport, the:

- risk was low
- profile was low
- benchmark QA level was basic
- achieved assessment is still in progress

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

- 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 QAAD assessment, still in progress
- 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. The QAAD will be updated as new alternative data sources are introduced into live production.

## 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 budgets<sup>1</sup>. 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. The CPIH was re-designated as a National Statistic on 31 July 2017. More details on CPIH can be found via the CPIH compendium.

### 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. For more information on how we assessed the overall assurance level, please see the [UK Statistics Authority Quality assurance matrix](#).

## 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.

### Breakdown of the four practice areas associated with data quality

The operational context and admin data collection practice area covers:

- the environment and processes for compiling the administrative data
- factors which affect data quality and cause bias
- safeguards which minimise the risks
- the role of performance measurements and targets; the potential for distortive effects

The communication with data partners practice area covers:

- collaborative relationships with data collectors, suppliers, IT specialists, policy and operational officials
- formal agreements detailing arrangements
- regular engagement with collectors, suppliers and users

The Quality Assurance (QA) principles, standards and checks by data suppliers practice area covers:

- data assurance arrangements in data collection and supply
- quality information about the data from suppliers
- role of operational inspection and internal/external audit in data assurance processes

The producers' QA investigations and documentations practice area covers:

- QA checks carried out by the statistics producer
- quality indicators for input data and output statistics
- strengths and limitations of the data in relation to use
- explanation for users about the data quality and impact on the statistics

## 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 (expect for rail fares which has been updated and provides weights information in February 2024).

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 below. 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.

### Benchmark assurance levels and assessment

For Autotrader, used cars, the benchmark risk assessment:

- for risk was low
- for profile was high
- overall was enhanced.

The justifications for this are:

- used cars have low weight contribution.
- using alternative data sources for used cars is methodological step change and there have been high levels of public engagement to this release.
- a data sharing agreement, a stable automated data feed, and regular meetings with the supplier are in place.

Overall, actual risk assessment is still progress.

For RDG LENNON, the benchmark risk assessment:

- for risk was low
- for profile was high
- overall was enhanced

The justifications for this are:

- rail fares have a low weight contribution
- high levels of public engagement have taken place prior to the release and rail fares will be the first alternative data source to go into live production
- the contract, service level agreement, and regular meetings with the supplier are in place

Overall, the actual risk assessment is still in progress.

For HHFCE, the benchmark risk assessment:

- for risk was high
- for profile was high
- overall was comprehensive

The justifications for this are:

- a complex data source is compiled from numerous data sources
- HHFCE are extensively used in CPIH, with no alternative data available
- regular communication with the supplier and some information on methodology is published

Overall, the actual risk assessment was enhanced: not achieved.

For TNS, the benchmark risk assessment:

- for risk was medium
- for profile was high
- overall was comprehensive

The justifications for this are:

- data accounts for high proportion of prices in CPIH and CPI
- a dedicated contract management branch assesses TNS's performance against contract
- sampling frame and design are set by prices, with quality checks carried out by both parties

Overall, the actual risk assessment was comprehensive: achieved.

For Value Office Agency (VOA) rental price data, the benchmark risk assessment:

- for risk was medium
- for profile was high
- overall was comprehensive

The justifications for this are:

- a relatively high weight in CPIH
- microdata is provided, allowing thorough quality assurance
- OOH costs are economically important, with high user interest in methodology

Overall, the actual risk assessment was comprehensive: achieved.

For LCF, the benchmark risk assessment:

- for risk was low
- for profile was high
- overall was enhanced

The justifications for this are:

- survey data which represents most items at lower-level aggregation
- collection, design and methodology are produced by the Office for National Statistics (ONS) and are well documented
- data are used widely in construction of economically important CPIH and CPI

Overall, the actual risk assessment was enhanced: achieved.

For Mintel, the benchmark risk assessment:

- for risk was medium
- for profile was medium
- overall was enhanced

The justifications for this are:

- although data collection is complex, the process and procedures are well documented
- a contract is in place with a designated contact
- data do not represent as broad a cross-section of the basket as other data sources

Overall, the actual risk assessment was comprehensive: achieved.

For Scottish Government rental data, the benchmark risk assessment:

- for risk was low
- for profile was medium
- overall was enhanced

For Northern Ireland Housing Executive rental data, the benchmark risk assessment:

- for risk was low
- for profile was medium
- overall was enhanced

The justifications for this are:

- a very low weight component of CPIH
- microdata is provided, allowing thorough quality assurance
- less user and media interest on devolved regions

Overall, the actual risk assessment was comprehensive: achieved

For Welsh Government, the benchmark risk assessment:

- for risk was low
- for profile was medium
- overall was enhanced

The justifications for this are:



- a very low weight component of CPIH
- microdata is provided, allowing thorough quality assurance
- less user and media interest on devolved regions

Overall, the actual risk assessment was comprehensive: achieved

For BEIS, the benchmark risk assessment:

- for risk was low
- for profile was low
- overall was basic

The justifications for this are:

- data sources are collected through a survey, with relatively low weight contribution
- methodology and quality assurance procedures are considered sufficient, and there is a dedicated BEIS contact
- limited, niche media interest in items

Overall, the actual risk assessment was basic to enhanced: achieved.

For Brochures, the benchmark risk assessment:

- for risk was low
- for profile was low
- overall was basic

The justifications for this are:

- a small to medium weight, with a number of mitigating factors that reduce the risk, for instance being collected in-house
- manually entered into prices system with robust quality assurance and validation
- little media interest in items

Overall, the actual risk assessment was basic: achieved.

For Consumer Intelligence, the benchmark risk assessment:

- for risk was low
- for profile was low
- overall was basic

The justifications for this are:

- a very low weight in CPIH, with clear contingency should data become unavailable
- a contract is in place with regular supplier meetings
- little media interest in item index

Overall, the actual risk assessment was basic: achieved.

For Department of Transport, the benchmark risk assessment:

- for risk was low
- for profile was low
- overall was high

The justifications for this are:

- it contributes a small to medium weight, with suitable alternatives if data unavailable
- data are sourced through email contact and imported directly into prices system
- the series is of limited media and user interest

Overall, the actual risk assessment is still in progress.

For Direct Contact, the benchmark risk assessment:

- for risk was low
- for profile was low
- overall was basic

The justifications for this are:

- a small to medium weight, with a number of mitigating factors that reduce the risk being collected in-house
- manually entered into prices system with robust quality assurance and validation
- little media interest in items

Overall, the actual risk assessment was basic: achieved.

For Glasses, the benchmark risk assessment:

- for risk was low to medium
- for profile was low
- overall was basic

The justifications for this are:

- complex data source is compiled from several different sources
- a relatively small weight contribution and alternative data sources being available
- the detail of quality assurance is not yet provided

Overall, the actual risk assessment was basic: incomplete.

For HESA, the benchmark risk assessment:

- for risk was low
- for profile was low
- overall was basic

The justifications for this are:

- a very low weight in CPIH, with clear contingency should data become unavailable
- data are sourced through email with no contractual agreement in place
- little media interest in item index

Overall, the actual risk assessment was basic: achieved.

For Home and Communities agency, the benchmark risk assessment:

- for risk was low
- for profile was low
- overall was basic

The justifications for this are:

- a very low weight in CPIH, with clear contingency should data become unavailable
- data are sourced through email with no contractual agreement in place
- little media interest in item index

Overall, the actual risk assessment was basic: achieved.

For IDBR, the benchmark risk assessment:

- for risk was low
- for profile was low
- overall was basic

The justifications for this are:

- a very low weight in CPIH, with clear contingency should data become unavailable
- the IDBR team is based within the ONS, so no contract is in place
- little media interest in item index

Overall, the actual risk assessment was basic: achieved.

For IPS, the benchmark risk assessment:

- for risk was low
- for profile was low
- overall was basic

The justifications for this are:

- a low but not insignificant weight in headline CPIH
- straightforward collection; primarily survey, supplemented by administrative
- methodology and quality assurance are well documented

Overall, the actual risk assessment was basic to enhanced: achieved.

For Kantar, the benchmark risk assessment:

- for risk was low
- for profile was low
- overall was basic

The justifications for this are:

- a very low weight in CPIH, with clear contingency should data become unavailable
- data are purchased annually with a dedicated contact provided
- little media interest in item index

Overall, the actual risk assessment was basic: achieved.

For Moneyfacts, the benchmark risk assessment:

- for risk was low
- for profile was low
- overall was basic

The justifications for this are:

- a very low weight in CPIH, with clear contingency should data become unavailable
- data are acquired through an annual magazine subscription
- little media interest in item index

Overall, the actual risk assessment was basic: achieved.

For Website, the benchmark risk assessment:

- for risk was low
- for profile was low
- overall was basic

The justifications for this are:

- a small to medium weight, with a number of mitigating factors that reduce the risk, for instance being collected in-house
- manually entered into prices system with robust quality assurance and validation
- little media interest in items

Overall, the actual risk assessment was basic: achieved.

## **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.

## 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 14% of the weight in the headline index. The private rental index accounts for 4% of the weight in the headline index, the majority of which will be due to England data.

Details on the quality assurance of administrative data can be found in our [Quality assurance of administrative data used in Private Rental Housing statistics](#).

## 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.

### Autotrader, used car data

#### Data usage

The data we receive from Autotrader is a point in time, snapshot of all live adverts across all vehicle types, taken daily. The dataset is rich with detailed attribute information, making it easier to define a unique product. As this is not an ecommerce platform but rather an online marketplace, we can only see listed prices in the data. The data are used to compile the second-hand cars index, thus elementary aggregates of the various car make indices, aggregated over the different fuel types (petrol, diesel, hybrid and electric), then into different age groups to construct the second-hand cars index. For more information, please see our [Using Auto Trader car listings data to transform consumer price statistics, UK methodology](#).

Risk: low

Used cars carries a low weight in the Consumer Prices Index including owner occupiers' housing costs (CPIH) and the Consumer Prices Index (CPI). Receiving daily data has enabled us to build contingency into the process, as we are able to iteratively run the indices as more data is received. This allows us to identify and resolve issues early in the production round and effectively manage the risks. We have established a good relationship with the Autotrader data team and our terms of engagement are also governed by a data sharing agreement.

Profile: high

Considerable public engagement and interest has been generated with regards to the transformation of consumer price statistics. Used cars data will be the second alternative data source to go into live production as part of a continuous programme of improvement. Using an alternative data source for our used cars index marks a significant improvement in our methodology.

Assessment: enhanced

Status: in progress

As a remedial action

- We will seek further information from Autotrader on data collection, methodology, and quality assurance procedures to allow us to make an assessment of the data source.



## 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.

## Rail Delivery Group

### Data usage

The Rail Delivery Group (RDG) produces data from the Latest Earnings Networked Nationally Over Night (LENNON) dataset for the Office for National Statistics (ONS) that include transaction level rail fares data, including expenditure and quantities, for rail journeys in Great Britain. These data are used to compile the rail fares indices, which are ranked by region and fare group (such as peak, off-peak, advance).

Risk: Low

Rail fares have a weight in headline the Consumer Prices Index including owner occupiers' housing costs (CPIH) and the Consumer Prices Index (CPI) of 0.9% and 1.1% respectively in 2023. Our relationship with RDG is governed by a legal contract with important performance indicators linked to quality assurance. Receiving daily data has enabled us to build contingency into the process, as we are able to repeatedly run the indices as more data are received. This allows us to identify and resolve issues early in the production round and effectively manage the risks.

Profile: High

The transformation of consumer price statistics has generated a lot of public engagement and interest. Rail fares will be the first alternative data source to go into live production as part of a continuous programme of improvement.

Assessment: In progress

Status: In progress

As RDG provides us with transaction level data, there are a limited number of transformations to be carried out on the data before it is sent to us, reducing the risk of error and bias within the data. Furthermore, the data were acquired based on detailed technical specification provided by the ONS. We are satisfied that our engagement with RDG has resulted in the acquisition of high-quality data. We are currently working with RDG to get the additional information required to complete this assessment.

Remedial actions:

1. We will seek further information from RDG on data collection, methodology, and quality assurance procedures to assess the data source.

## Scottish government rental data

Scottish government data are used to produce the rental price series for Scotland in the private rental price index, and the OOH component of CPIH. Scottish government data are also used to produce strata weights for the Scotland stratum of the OOH component in CPIH. This stratum has a weight of around 1% in CPIH. Scottish government data are likely to represent a small proportion of the approximately 4% weight for the private rents index.

Details on the quality assurance of administrative data can be found in our [Quality assurance of administrative data used in Private Rental Housing statistics](#).

## Welsh government

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 around 0.6% in CPIH. Welsh government data are likely to represent a small proportion of the approximately 4% weight for the private rents index.

Details on the quality assurance of administrative data can be found in our [Quality assurance of administrative data used in Private Rental Housing statistics](#).

## Northern Ireland Housing Executive

Northern Ireland Housing Executive data are used to produce the rental price series for Northern Ireland in the Price Index of Private Rents, and the OOH component of CPIH. Northern Ireland data are also used to produce strata weights for the Northern Ireland stratum of the OOH component in CPIH. This stratum has a weight of around 0.4% in CPIH. Northern Ireland data are likely to represent a small proportion of the approximately 2% weight for the Price Index of Private Rents.

Details on the quality assurance of administrative data can be found in our [Quality assurance of administrative data used in the Price Index of Private Rents](#).

## 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 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.

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, with details on what actions we intend to take to rectify them.

### **HHFCE**

HHFCE are in the process of completing a QAAD assessment of their data sources.

### **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**

We plan to establish better lines of communication with Glasses by seeking a dedicated point of contact within the company. We will continue to request information on quality assurance procedures from Glasses.

### **Consumer Intelligence**

We will clarify contact details for Consumer Intelligence. We will seek further quality assurance information from Consumer Intelligence.

### **Moneyfacts**

We plan to establish a dedicated contact for Moneyfacts and keep this up to date. We will seek further quality assurance information from Moneyfacts.

## Various

We will set up better communication mechanisms and establish firmer data delivery agreements with various data sources.

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.

## . Annex A: Assessment of data sources

### 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

The quality assurance of the Valuation Office Agency rental price data can be found in the [Price Index of Private Rents Quality Assurance of Administrative Data](#).

## 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**

The quality assurance of the Scottish Government rental price data can be found in the [Price Index of Private Rents Quality Assurance of Administrative Data](#).

## **8. Welsh government**

The quality assurance of the Welsh Government rental price data can be found in the [Price Index of Private Rents Quality Assurance of Administrative Data](#).

## **9. Northern Ireland Housing Executive**

The quality assurance of the Northern Ireland Housing Executive rental price data can be found in the [Price Index of Private Rents Quality Assurance of Administrative Data](#).

## **10. 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



## 11. 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

## 12. 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

## **13. 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.

## **14. 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.

## **15. Rail Delivery Group (RDG)**

### **Practice area 1: Operational context and admin data collection**

The Latest Earnings Networked Nationally Over Night (LENNON) dataset captures at least 90% of all rail fares sold by train operating companies at a station or through a third-party application, among other options. This excludes rail fares for journeys that exclusively take place within the London Underground system. Daily data processed by RDG are received by the Office for National Statistics (ONS) through an automated feed.

### **Practice area 2: Communication with data supplier partners**

The ONS holds regular meetings with RDG at least once every month. In these meetings we discuss any data quality issues raised as part of our internal quality assurance processes. Where applicable, we also discuss any upcoming changes to the data feed, carefully assessing the impact of those changes and sequencing those changes for seamless integration. There is a legal contract between the ONS and RDG that ensures any issues with data delivery are resolved in a timely manner.

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

RDG stores and maintains the LENNON dataset using data management best practice to reduce the risk of recording error. As this dataset underpins a core part of their business model, RDG has built in multiple layers of validation to ensure accurate recording of transactions.

### **Practice area 4: Producers' quality assurance investigations and documentation**

Internally the ONS carries out a series of quality assurance procedures that include:

- basic checks as the data are received to ensure good coverage of all main variables and correct file formats

- further validation of completeness on a weekly basis by measuring the proportion of null values within the data; these must fall within a set of predefined thresholds, which are actively monitored to maintain the high quality of our outputs
- expenditure patterns are tracked and any irregularities are flagged for further investigation
- indices are reviewed on a weekly basis to identify and resolve issues early in the production round

### **Missing data or imputation**

Because of having almost complete coverage of the rail fares industry and the high volume of daily transaction level data which contain both price and quantities, imputation is not used in producing these indices.

Missing data are handled in line with a Service Level Agreement (SLA) between the ONS and RDG. Typically, any issues with missing data are resolved within 24 hours.

We have also set out a high-level contingency plan for consumer price inflation statistics if new large sources, such as rail fares data, are unavailable or not of sufficient quality for inclusion in the monthly publication round.

Revisions policy

Please refer to the ONS's [Revisions policy for consumer price inflation statistics article](#).

## **16. 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.

## 17. 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 analyst<sup>1</sup> and the signed printout of the spreadsheet is put on the working file with the Price Data.

## **18. Autotrader used cars data**

### **Practice area 1: operational context and admin data collection**

Autotrader is the UK's largest online marketplace for the automotive industry where car dealers and private individuals can advertise any vehicle for sale. It is worth noting that, Autotrader isn't an ecommerce site that would handle the sales transactions of any vehicles advertised on their site. Therefore, the price recorded in the data refers to the listing price. Data is ingested into the Office for National Statistics (ONS) on a daily basis through an automated pipeline.

### **Practice area 2: communication with data supplier partners**

The ONS holds regular meetings with Autotrader at least once every month, where we catch up with our data contact. In these meetings, we discuss any data quality issues raised as part of our internal quality assurance processes. Where applicable, we also discuss any upcoming changes to the data feed, carefully assessing the impact of those changes and sequencing those changes for seamless integration. Terms of engagement with Autotrader are governed by a data sharing agreement (DSA), which sets out the format and quality of the data we expect to receive. Finally, we also have an escalation contact in case an issue cannot be resolved using our usual process.

### **Practice area 3: quality assurance principles, standards and checks by data suppliers**

As we receive a very granular cut of data, the risk of data quality being affected by downstream processes is limited. Several workshops have been held for Autotrader to better understand how we use their data, which has been fed into the quality checks they do on the data before sending it to the ONS.

### **Practice area 4: producers' quality assurance investigations and documentation**

Internally, the ONS carries out a series of quality assurance procedures that include:

- basic checks as the data are ingested to ensure good coverage of all key variables and correct file formats.

- further validation on a weekly basis, to measure of completeness by measuring the proportion of null values within the data; t. These must fall within a set of predefined thresholds and the thresholds are actively monitored to maintain the high quality of our outputs.
- expenditure patterns are being tracked, and any anomalies being flagged for further investigation.
- indices are being reviewed on a weekly basis to identify and resolve issues early in the production round.

## Missing data or imputation

Autotrader being the UK's the largest online marketplace for used cars means we have very good coverage of the used cars market, the gap being filled by other relatively smaller players in the used cars market. The Autotrader data is high volume and high frequency and as a result,, imputation is not used in producing these indices.

Missing data are handled in line with the DSA between the ONS and Autotrader.

We have also set out a high-level contingency plan for consumer price inflation statistics if new large sources, such as the used cars data, are unavailable or not of sufficient quality for inclusion in the monthly publication round.

## Revisions policy

Please refer to the ONS's Revisions policy for consumer price inflation statistics article.

## 19. 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.

## **20. 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.

### 21. 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 177/2008 of 20 February 2008 establishing a common framework for business registers for statistical purposes
- Council Regulation (EEC) No 696/93 on statistical units for the observation and analysis of the production system in the Community; and 6
- Commission Regulation (EC) No 250/2009 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.



## 22. 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.

## **23. 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.

## **24. 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.

## **25. 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

## **. Annex B: ONS data checking and validation**

For information relating to data checking and validation of our price data, please see our [Consumer Prices Indices Technical Manual](#).

## **. Annex C: 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

## . Annex D: Flow diagrams of quality assurance processes

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