Guidelines for Measuring Statistical Output Quality

Version 4.1

September 2013

Copyright

© Crown copyright 2013
You may use or re-use this information (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence. To view this licence, visit www.nationalarchives.gov.uk/doc/open-government-licence/ or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gsi.gov.uk.

This document is also available on our website at www.ons.gov.uk.

Contact details

For enquiries about this publication, contact:
Quality Centre
Email: ONS.Quality.Centre@ons.gsi.gov.uk

Preface

Quality is central to the statistics that we all produce across the GSS. As National Statistician**,** it is my responsibility to ensure that we deliver statistics that are of high quality and integrity, are accessible and well explained and win the trust and confidence of our users.

A great deal of research has gone into the preparation and updating of these guidelines to ensure that they promote a common approach and outline best practice for measuring and reporting on quality across the Government Statistical Service (GSS). The guidance is a tool that will help us to ensure that the quality requirements of the Code of Practice for Official Statistics are met and that the GSS vision; 'statistics produced by a collaborative community, used in decision making and debate' is realised. It will also provide consistency in information about our statistics that will allow users to judge for themselves the quality and appropriate uses of the data in accordance with their needs.

Using these guidelines when planning and producing statistics and compiling statistical reports and publications, will help ensure their quality. I know this is important to us all.

Jil Matheson
National Statistician
September 2013

Contents

Section A: Introduction 5

A1. Background 6

A2. Aim and purpose of the guidelines 6

A3. What is ‘quality’? 7

A4. Key Quality Measures 10

Section B: Quality measurement guidelines 11

B1. Specifying user needs 14

B2. Design and build 19

B3. Collection 25

B4. Processing 35

B5. Analysis 39

B6. Dissemination 67

Section A: Introduction

The Guidelines for Measuring Statistical Output Quality (Version 4.1) provide a checklist of quality measures and indicators for use when measuring and reporting on the quality of statistical outputs. They are not a requirement but represent good practice for measuring quality throughout the statistical production process.

This version of the guidelines replaces Version 3.1 (2007). The quality measures and indicators have been updated to serve a wider range of statistical data, rather than predominantly focussing on survey data. Thus, the guidelines address the growing use of administrative data in producing statistical outputs.

The purpose of this document is to promote a standardised approach to measuring and reporting on quality across the Government Statistical Service (GSS). Many of the measures and indicators in the guidelines will be familiar to government statisticians – for example, standard errors and response rates. Others will be less familiar as they are less commonly used.

Although the guidelines are primarily aimed at producers of official statistics, they can be used by anyone wanting to report on the quality of statistical outputs.

This document consists of two parts. Section A presents background information on the guidelines. Section B provides the list of quality measures and indicators for consideration when producing statistical outputs.

A1. Background

The GSS is committed to providing users with information on the quality and reliability of its statistical outputs, along with the methods that have been used to produce them. This commitment is expressed in Principles 4 and 8 of the Code of Practice for Official Statistics:

> **Principle 4, Practice 2**
> *'Ensure that official statistics are produced to a level of quality that meets users' needs, and that users are informed about the quality of statistical outputs, including estimates of the main sources of bias and other errors, and other aspects of the European Statistical System definition of quality.'*
>
> **Principle 4, Practice 4**
> *'Publish quality guidelines and ensure that staff are suitably trained in quality management.'*
>
> **Principle 8, Practice 1**
> *'Provide information on the quality and reliability of statistics in relation to the range of potential uses, and on methods, procedures, and classifications.'*

The guidelines aim to aid producers in fulfilling this commitment and can also be used as a building block to comply with the National Statistician's Guidance on Quality, Methods and Harmonisation, which states that:

> *'Each organisation should have a policy which states where and which quality output measures will be reported. For example, the organisational policy may state that all first releases will include a core set of quality measures or include a web link to quality information; while statistical press releases would contain this information in their 'Notes to editors'. All key statistical outputs should have basic quality information as the minimum. The policy may also state that for each statistical product a reference report is produced which contains measures that do not change from one release to another.'*

A2. Aim and purpose of the guidelines

The overall aim of the guidelines is to outline best practice for measuring and reporting on the statistical quality of GSS outputs. In particular, the emphasis is upon helping users to understand:

- the context in which the data were collected, processed and analysed;
- the extent to which the data relate to potential uses;
- methods adopted and limitations they impose;
- the reliability of the figures;
- the way they relate to other available data on the same subject; and
- the availability of the statistical output.

The measures and indicators can also be used by producers of official statistics to monitor data quality for the purpose of continuous improvement.

A3. What is ‘quality’?

The word ‘quality’ has many different meanings, depending on the context in which it is used. The quality of statistical outputs is most usefully defined in terms of how well outputs meet user needs, or whether they are ‘fit for purpose’. This definition is a relative one, allowing for various perspectives on what constitutes quality, depending on the intended uses of the outputs.

In order to enable users to determine whether outputs meet their needs, it is recommended that output producers report quality in terms of the five quality dimensions of the European Statistical System (ESS), which are: Relevance; Accuracy and Reliability; Timeliness and Punctuality; Accessibility and Clarity; and Coherence and Comparability. The quality measures and indicators in Section B of the guidelines have been developed around these five dimensions. A good summary of quality should contain quality measures and indicators for each of the five ESS quality dimensions and should also include elements of the other considerations listed within table A.1.

Quality measure or quality indicator?

Quality measures are defined as those items that directly measure a particular aspect of quality. For example, the time lag from the reference date to the release of the output is a direct measure. However, in practice, many quality measures can be difficult or costly to calculate. Instead we can use quality indicators to give insight into quality.

Quality indicators usually consist of information that is a by-product of the statistical process. They do not measure quality directly but can provide enough information to provide an insight into quality. For example, in the case of accuracy it is almost impossible to measure non-response bias as the characteristics of those who do not respond can be difficult to ascertain. In this instance, response rates are a suitable quality indicator that may be used to give an insight into the possible extent of non-response bias.

The guidelines include both quality measures and quality indicators, which can either supplement or act as substitutes for the desired quality measures.

Table A.1 Dimensions of quality

Definition	Key components
1. RELEVANCE	
The degree to which statistics meet current and potential needs of the users.	Any assessment of relevance needs to consider: • who are the current and potential users of the statistics; • what are their needs; and • how well the output meets these needs.
2. ACCURACY AND RELIABILITY	
The closeness between an estimated result and the (unknown) true value.	Accuracy can be split into sampling error and non-sampling error, where non-sampling error includes: • coverage error; • non-response error; • measurement error; • processing error; and • model assumption error.
3. TIMELINESS AND PUNCTUALITY	
Timeliness refers to the lapse of time between the period to which the data refer and publication of the estimate. Punctuality refers to the time lag between the actual and planned dates of publication.	An assessment of timeliness and punctuality should consider the following: • production time; • frequency of release; and • punctuality of release.
4. ACCESSIBILITY AND CLARITY	
Accessibility is the ease with which users are able to access the data. It also relates to the format(s) in which the data are available and the availability of supporting information. Clarity is the extent to which easily comprehensible metadata are available, where these metadata are necessary to give a full understanding of the statistical data.	Specific areas where accessibility and clarity may be addressed include: • needs of analysts; • assistance to locate information; • clarity; and • dissemination format.
5. COHERENCE AND COMPARABILITY	
Coherence is the degree to which the statistical processes, by which two or more outputs are generated, use the same concepts and harmonised methods. Comparability is the degree to which data can be compared over time, region or other domain.	Coherence should be addressed in terms of coherence between: • data produced at different frequencies; • other statistics in the same socio-economic domain; and • sources and outputs. Comparability should be addressed in terms of comparability over: • time; • spatial domains (eg sub-national, national, international); and • domain or sub-population (eg industrial sector, household type).

Other Considerations

Beyond the five dimensions of statistical output quality outlined in Table A.1, the European Statistical System Handbook for Quality Reports also identifies other considerations which are relevant to measuring and reporting on quality. Further information from the handbook is included below.

Trade-offs between output quality components: The dimensions of statistical output quality listed in Table A.1 are not mutually exclusive; there are relationships between the factors and there are instances where improvements in one dimension would lead to deterioration in another dimension. For example, increasing the timeliness of the output may mean a deterioration in the accuracy as less data has been received and the producer has been able to make fewer checks on the raw data. Trade-offs that have to be made in these circumstances should be communicated to users, along with the reasons that those decisions were made.

Other considerations cover those aspects of process quality that are not readily reported under any of the output components:

- **Assessment of user needs and perceptions:** Users should be the starting point for quality considerations and so information regarding their needs and perceptions should be obtained for all output components and information should be provided on how these have been addressed.
- **Cost, performance and respondent burden:** These are important process quality components that are not readily covered by the output quality dimensions. There are invariably trade-offs required between all of the output quality components and cost, performance and response burden. Details should be provided on how these considerations have impacted the output.
- **Confidentiality, transparency and security:** Likewise, these are process quality components not covered elsewhere but which can have an impact on the output produced.

Further information about each of these considerations is available within the ESS Handbook for Quality Reports. An updated version is due to be published in late 2013, although many of the updates have already been incorporated within this document.

A4. Key Quality Measures

In previous versions of these guidelines, a few quality measures have been highlighted as those that are key to providing users with an overall summary of output quality. However, with developments in a number of areas, including the increasing use of administrative data and more complex mixed source (survey and administrative data) combinations of data used to produce statistical outputs, a single list of a few Key Quality Measures (KQMs) is no longer feasible.

The quality measures and indicators throughout this document have been developed primarily for statistical producers, who should select the quality measures and indicators most relevant to their output(s). These measures can be used to provide management information, to monitor performance and to assess any quality improvements in statistical outputs.

The quality measures and indicators should also be used for quality reporting, thus benefiting users of the statistical outputs. They provide the user with an indication of the quality of the output, and an awareness of how the data have been used in the production of the output.

Various quality measures and indicators are included throughout this document. As a minimum requirement, producers of statistics should comment on all of the ESS quality dimensions and wider considerations (see Table A.1). All quality measures and indicators that would be relevant to their output should be included in order to aid user understanding and enable users to determine whether the output(s) meet their needs. In previous versions of these guidelines, there was a separate section of quality measures and indicators specifically for administrative data. These have now been incorporated into the main sections which have been made more generic/applicable to survey and administrative data based outputs and have been labelled as such (see Section B).

Further information on quality measures and indicators can be found in the relevant sections.

In recent years**,** there has been a marked increase in the use of administrative data in the production of statistics, within the UK and across Europe. In some cases**,** the administrative data is being used directly to produce the statistical outputs**,** but in many cases**,** the administrative data are being combined with survey data. With this increasing use of administrative data comes increasingly complex ways of using the data. In these contexts, some of the simple input and process indicators (as included throughout this document) may be less informative of output quality. Since sampling theory typically does not apply to administrative data, the assessment of the accuracy of these mixed source statistics is a challenge. The ESSnet Admin Data has provided some further guidance on this topic which is available on the ESSnet Admin Data Information Centre (see specifically Deliverables 2011/6.3 and 6.5[1]).

Further guidance on assessing the quality of the administrative data itself has also been developed. This can be useful in informing the quality of the statistical output, but is largely out of scope of this document. Further information on assessing the administrative sources and the quality of administrative data is available from the ESSnet Admin Data Information Centre (see specifically Deliverables 2011/2.2 and 2.4[2]) and the BLUE-ETS project (see specifically Deliverable 4.3[3]).

[1] http://essnet.admindata.eu/WorkPackage?objectId=4257
[2] http://essnet.admindata.eu/WorkPackage/ShowAllDocuments?objectId=4252
[3] http://www.blue-ets.istat.it/index.php?id=7

Section B: Quality measurement guidelines

A checklist of items to consider when reporting on the quality of statistical outputs is presented over the following pages. It is not the intention that all quality measures should be addressed for all outputs. Instead, the output producer is encouraged to select those quality measures and indicators that together provide users with an indication of the overall strengths, limitations and appropriate uses of a given dataset.

The quality measures and indicators are grouped together into stages of the statistical production process: specifying user needs; design and build; collection; processing; analysis; and dissemination.

These stages broadly align to the list of statistical processes outlined in the Generic Statistical Business Process Model (GSBPM), which was produced by the United Nations Economic Commission for Europe (UNECE), the Organisation for Economic Co-operation and Development (OECD) and Eurostat[4].

The GSBPM is intended to apply to all activities undertaken by producers of official statistics, at both the national and international levels, which result in data outputs. It is designed to be independent of the data source, so it can be used for the description and quality assessment of processes based on surveys, censuses, administrative records, and other non-statistical or mixed sources.

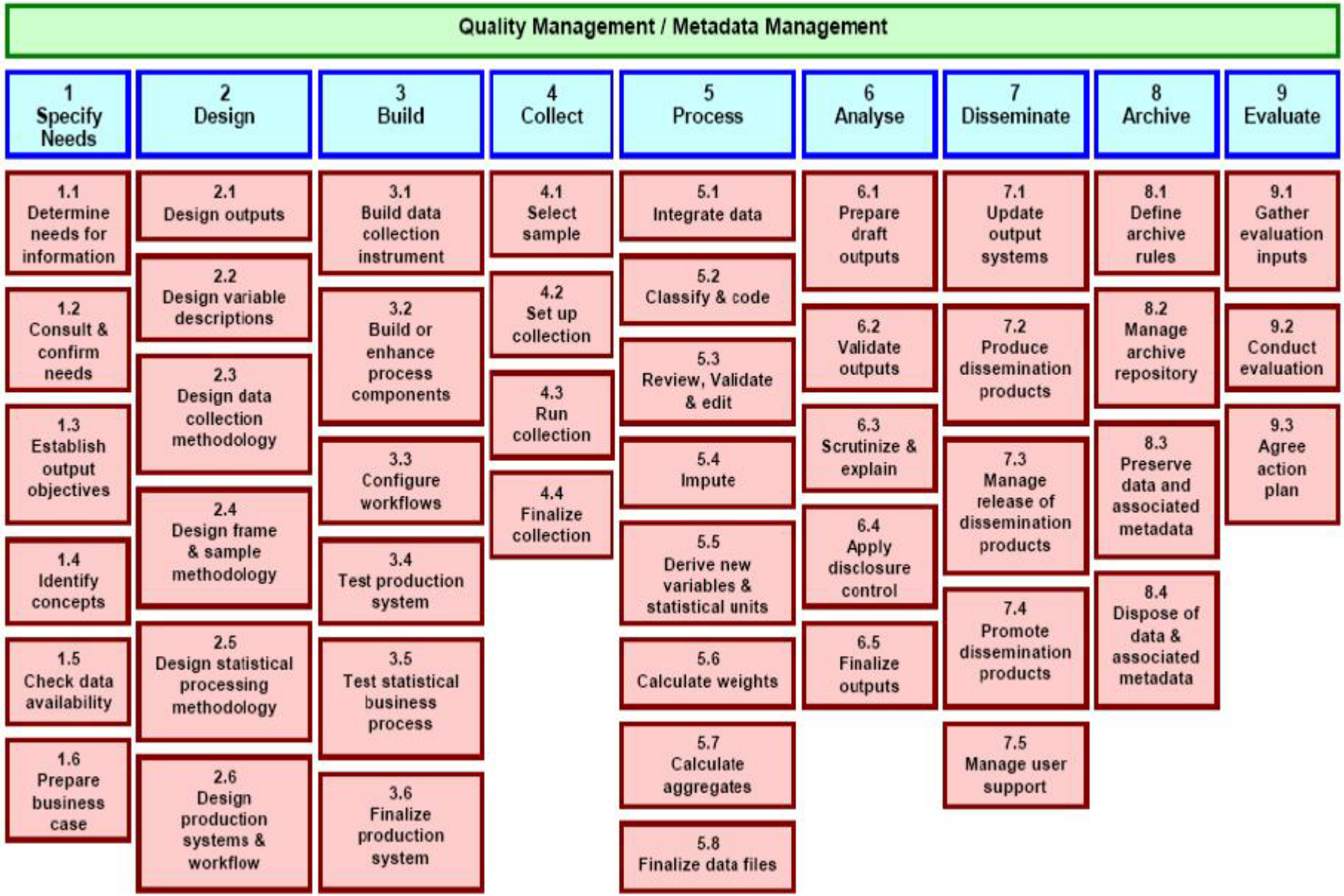
The GSBPM comprises four levels:

- Level 0, the statistical business process;
- Level 1, the nine phases of the statistical business process;
- Level 2, the sub-processes within each phase;
- Level 3, a description of those sub-processes.

Figure B.1 displays the GSBPM at Level 2.

[4] More information on the GSBPM can be found here:
http://www1.unece.org/stat/platform/display/metis/Generic+Statistical+Business+Process+Model

Figure B.1 The Generic Statistical Business Process Model (GSBPM)



The GSBPM categorisation shown in Figure B.1 is too detailed for classifying the quality measures in these guidelines. However, the six categories we have chosen for the quality measures roughly correspond to the phases of the statistical business process in Level 1 of the GSBPM, as shown in Figure B.1.

In addition to the categorisation by stages of the statistical production process, the quality measures have also been grouped into the ESS quality dimensions: Relevance; Accuracy and Reliability; Timeliness and Punctuality; Accessibility and Clarity; and Coherence and Comparability. More information on these is included in Table A.1.

The tables in this section contain quantitative and qualitative quality measures and indicators, together with:

- descriptions of each measure/indicator and notes on use;
- an example of the type of information you may want to record when addressing each measure (or in some cases for the quantitative measures, the formulae used to calculate the measure); and
- an identification of measures and indicators that are relevant to outputs involving either survey data, administrative data or both (see Table B.1).

The type of data used in the production of the statistical output impacts which quality measures and indicators can be used. It is envisaged that some quality measures and indicators will be applicable only to outputs involving survey data (for example, B2.7: Describe the sample design). These quality measures and indicators are identified with a ▲ . Other quality measures and indicators will be applicable only to outputs involving administrative data (for example, B3.1: For outputs that require the use of administrative sources, describe the timeliness and punctuality of the data received). These types of quality measures and indicators are identified with a ● . Alternatively, some quality measures and indicators will be applicable to outputs involving either survey data, administrative data or both (for example, B1.1: Describe and classify key users and uses of the statistical output). These quality measures and indictors are identified with a ■ .

Table B.1 Data Type

Data Type	Description
Outputs involving survey data ▲	Data derived from a survey provides information on the characteristics of a given population. This is achieved by collecting data from a sample of that population and estimating their characteristics through the systematic use of statistical methodology.
Outputs involving administrative data ●	Administrative sources are sources containing information that is not primarily collected for statistical purposes but is subsequently used to produce statistics.
Outputs involving either survey data, administrative data or both ■	Relevant to outputs derived from survey and/or administrative data.

Please note: efforts have been made to include examples from as wide a range of statistical outputs and contexts as possible. Where the examples have been taken from ONS quality documentation, those documents have been referred to specifically and are available on the ONS website here: http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-information/index.html

If you have any examples or quality measures and indicators that could be incorporated within these guidelines, please contact ONS Quality Centre (ONS.quality.centre@ons.gsi.gov.uk).

B1. Specifying user needs

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B1.1 Describe and classify key users and uses of the statistical output. *(Relevance)* ■	This information can be obtained from initial requests to produce the statistical output, user feedback, or from information on the users of previous similar outputs. The users are classified according to their use of the output and the type of agency they are affiliated to (for example, institutions; international organisations; researchers and students; and businesses).	Workforce Jobs (WFJ) estimates and data produced for the quarterly publication are used across government, business and academia, and feed into a number of wider publications and outputs. Some government departments use the total figures to facilitate policy making, whereas others use specific components of the published data. Some of the main users are: • HM Treasury - estimates are used to assess labour market conditions and are an important indicator in the context of macroeconomic assessment; • Bank of England (BoE) - the BoE Monetary Policy Committee is a user of WFJ estimates. The BoE use the estimates as an indicator for setting interest rates; • Department for Education (DfE) - DfE use WFJ to inform and monitor policy making such as the success of their Welfare to Work programme; • Eurostat - estimates of Employee Jobs (EJ) are supplied directly to Eurostat. This is a legal requirement under the Short Term Indicators Regulation (STIR); and • Devolved Administrations – Scottish and Welsh Governments. *Source: Workforce Jobs, Quality and Methodology Information, Office for National Statistics, 2013*
B1.2 Where possible, describe how the data relate to the needs of users. *(Relevance)* ■	This indicator captures how well the data support users' needs. This information can be gathered from user satisfaction surveys and feedback.	When the Index of Services (IoS) project was initiated in early 1999, a user requirement was drawn up outlining the key users of the IoS including the Bank of England (BoE), HM Treasury (HMT), Department of Business, Innovation and Skills (BIS), along with their needs and requirements. As at February 2009, 14 of the 18 user requirements have been met. For example, for the IoS to cover the entire service sector, both public and private and value added weights need to be used, as in the national accounts. A programme of work is in place to deliver the remaining four requirements and users are kept informed of progress on a regular basis. *Source: Index of Services (IoS), Summary Quality Report, Office for National Statistics, 2009*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B1.3 Identify methods of user engagement. *(Relevance)* ■	This indicator should provide information on the different user engagement strategies that are in place for the output, to demonstrate the level of communication that takes place when producing that output. Examples of methods of user engagement could include: Statsusernet (discussion forum); output user groups; conferences/seminars/roadshows/focus groups; social media; communication via ad hoc requests; dealing with complaints; email / newsletters; face to face meetings; consultations and surveys; and TV/radio interviews.	Some of the key users of the Public Sector Finances (PSF) and Government Deficit and Debt under the Maastricht Treaty statistical bulletins are UK government organisations, such as: Office for Budget Responsibility (OBR) and the Bank of England (BoE); and international bodies, such as: Eurostat and the International Monetary Fund (IMF). ONS and Her Majesty's Treasury (HMT) are in regular dialogue with these users via a range of different UK and international working groups. The use made of the statistical bulletins is much wider than government. Extensive use is made of the bulletins by City analysts, rating agencies, research institutes and academics to inform their understanding of the state of the UK public sector finances and any related risks or opportunities. A still broader range of users are the media, academia and the general public who refer to the bulletins to inform themselves about the public sector finances and in particular levels of public expenditure, revenue, debt and borrowing. Users are consulted as to their views of the statistical bulletins and how well it meets their needs. These consultations take the form of both a web survey and workshop events which are open to the public. *Source: Public Sector Finances (PSF), Quality and Methodology Information, Office for National Statistics, 2012*
B1.4 Describe any recent user feedback or results of user satisfaction assessments and actions that have been taken as a result. *(Relevance)* ■	The main results of user satisfaction assessments should be reported, giving priority to the results for the most important groups of users. This indicator should also record how the results of customer feedback and satisfaction surveys are translated into concrete actions to improve the relevance of outputs, for example, changes to how concepts are operationalised as a result of user feedback on lack of relevance for their needs.	Feedback was sought from users of "Deaths involving *Clostridium difficile* in England and Wales" in the 2009 bulletin and responses were received from key stakeholders including Department of Health (DH) and National Health Service (NHS) Primary Care Trusts. Comments were made on the presentation of the bulletin and users raised concerns about the timing of the release and asked whether quarterly data could be made available. Changes to the format of the bulletin were made in response to these comments and figures for deaths registered in each quarter were added. It was not been possible to alter the timeliness of the release due to the requirement for the publication of the provisional annual death registrations data prior to publication. *Source: Deaths involving Clostridium Difficile, Quality and Methodology Information, Office for National Statistics, 2012*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B1.5 Describe the sources of the data. *(Relevance)* ■	This should include the known purpose of the data collection and known merits and shortcomings of the data. The information will help users to assess whether the output is relevant and of sufficient quality for their uses.	The three data sources used to calculate internal migration estimates are administrative data sources. This means that their use as a statistical data source is secondary and they were not designed for this purpose. Overall, the accuracy of the NHS based data depends on patients re-registering with a new doctor when they change residence or informing their current GP of any change in residential address. It is known that re-registration patterns vary by sex and age group; this is dealt with in more detail in the following reports: Report on Research into Revising Internal Migration Estimates and Update following Further Investigation into Revising Internal Migration Estimates. Research has been undertaken to assess the feasibility and benefits of adjusting the estimates to account for variations but no suitable data source currently exists that would facilitate any improvement to the current methodology. Data from the National Health Service Central Register (NHSCR) system are limited to movements between the former health authorities. These former health authorities no longer exist as administrative entities but continue to be used due to technical constraints of the NHSCR processing system. Patient Register Data Service (PRDS) data are also inadequate as a stand alone source as a number of within-year moves are not captured, including: 1. Migrant babies aged less than one year. 2. New non-birth registrations eg ex-armed forces personnel and international in-migrants that join the NHS and then move within the same year. 3. People who move during one year but then leave the NHS register before the end of the second year eg the deceased, new armed forces personnel and international out-migrants. Limitations of estimates derived from a combination of NHSCR and PRDS are: 1. There is variation in the delay between a person moving and registering with a new doctor. 2. Some moves may not result in a GP re-registration and therefore will not be recorded. 3. Individuals may move and re-register more than once in a single year but remain within the same health authority. Despite some limitations, research has shown that there is no one other available source that has as good coverage and quality as the combination of PRDS, NHSCR and Higher Education Statistics Agency (HESA) data. *Source: Internal Migration Estimates (IME), Quality and Methodology Information, Office for National Statistics, 2011*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B1.6 Describe key statistical concepts. *(Relevance)* ■	This should include descriptions of the statistical measure, the population, variables, units, domains and time reference. This information gives users an understanding of the relevance of the output to their needs, for example whether the output covers their required population or time period.	The Labour Force Survey (LFS) is a legal requirement of the European Union. The definitions of the main economic activity groups - the economically active, comprising the employed and the unemployed, and the economically inactive - which are used in the LFS, are standard ILO definitions. The economically active are defined as those aged 16 or over, who are either employed or unemployed in the survey reference week. The employed are defined as those aged 16 or over, who are in employment if they did at least one hour of work in the reference week (as an employee, as self-employed, as unpaid workers in a family business, or as participants in government-supported training schemes), and those who had a job that they were temporarily away from (for example, if they are on holiday). The unemployed are defined as those aged 16 or over, who are without work, have actively sought work in the last four weeks and are available to start work in the next two weeks; or are out of work but have found a job and are waiting to start it within the next two weeks. The economically inactive are defined as those aged 16 or over who are neither in employment nor unemployed. This group includes, for example, all those who are looking after a home or family, have a long-term illness or disability which prevents them working, or are retired. Unpaid family workers also belong to the employed category, and are those who are doing unpaid work in a family business. *Source: Labour Force Survey (LFS), Quality and Methodology Information, Office for National Statistics, 2009*
B1.7 Describe any gaps between measured statistical concept and the concept(s) users require. *(Relevance)* ■	The gap between the user's concept of interest and the measured statistical concept should be assessed and described. Reasons for this should also be provided, where possible. For example, gaps between what is measured and what users require may be due to definitional variations or differences in the way that concepts are measured.	Since Annual Survey of Hours and Earnings (ASHE) is a survey of employee jobs, it does not cover the self-employed or any jobs within the armed forces. Given the survey reference date in April, the survey does not fully cover certain types of seasonal work, for example, employees taken on for only summer or winter work. *Source: Annual Survey of Hours and Earnings (ASHE), Summary Quality Report, Office for National Statistics, 2011*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B1.8 Describe plans for meeting needs arising from lack of completeness. *(Relevance)* ■	This allows users to assess whether plans to ensure that outputs are complete, in terms of coverage and detail, are adequate for their needs of the output.	ONS has consulted with users of Internet Access statistics and resulting from this, new user interest has emerged in information about how many adults have in the past been Internet users, but who have not used the Internet recently. ONS is planning to address this by publishing new estimates of adults who have used the Internet, broken down into those who had used it within the three last months, and those who last used it more than three months ago. These new estimates will be added to the quarterly publication from 2012 Q2. This will be possible because the Internet use question 'when did you last use the Internet?' has response options 'Within the last three months', 'Between three months and a year ago' and 'More than one year ago'. Responses for 'Between three months and a year ago' and 'More than one year ago' will be aggregated to provide the new estimates of adults who last used the Internet more than three months ago. *Source: Internet Access - Quarterly Update, Quality and Methodology Information, Office for National Statistics, 2012*

Beyond the five dimensions of statistical output quality outlined in Table A1, the European Statistical System Handbook for Quality Reports also identifies other considerations which are relevant to measuring and reporting on quality. Further information is included on page 9.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

B2. Design and build

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B2.1 Describe the sampling frame. *(Accuracy and Reliability)* ▲	This description is to inform users how the sampling frame was constructed, and whether it is current or out of date. Other information, such as the variables contained within the sampling frame, what uses are made of it, who has ownership and whether there are special requirements for access, can be commented on to provide further detail. Deficiencies in the frame or its metrics, eg size and coverage of the population, should also be considered.	The Inter-Departmental Business Register (IDBR) is the sampling frame for surveys of businesses carried out by ONS and by other government departments. It is also a key data source for analysis of business activity. The main administrative sources for the IDBR are: Value Added Tax (VAT) trader and Pay As You Earn (PAYE) employer information supplied to ONS by HM Revenue & Customs (HMRC) under the Value Added Tax Act 1994 for VAT traders and the Finance Act 1969 for PAYE employers; details of incorporated businesses are also passed to ONS by Companies House. ONS survey data and survey information from the Department of Finance and Personnel, Northern Ireland (DEFPNI) and the Department for Environment, Food and Rural Affairs (DEFRA) farms register provide additional information. The IDBR combines this survey data in a statistical register comprising over two million enterprises. These comprehensive administrative sources combined with the survey data contribute to the coverage on the IDBR, which is one of its main strengths. The IDBR's coverage is very good, representing around 99% of economic activity; it covers businesses in all parts of the economy, missing some very small businesses operating without VAT or PAYE schemes (self-employed and those with low turnover and without employees) and some non-profit organisations. Variables held on the IDBR include; business name, address including postcode, Standard Industrial Classification (UK SIC 2007 and UK SIC 2003), employment and employees, turnover, legal status (company, sole proprietor, partnership, public corporation/nationalised body, local authority or non-profit body), enterprise group links, country of ownership and company number. *Source: Introduction to the Inter-Departmental Business Register (IDBR), Office for National Statistics website*
B2.2 How is the frame maintained and when was it last updated to take account of changes in the survey population and/or changes in classifications? *(Accuracy and Reliability)* ▲	Populations are rarely constant and it is important that sampling frames are updated with information on births, deaths and any changes in classification to units in the population. Reporting on updating practices gives an indication to users of the quality of the sampling frame.	Maintaining the accuracy of the data in the Postcode Address File (PAF) is the Address Management Unit's number one priority. Royal Mail's 65,000 postmen and women capture thousands of address changes every day that are then used to update PAF on a quarterly basis. ONS uses the PAF as the sampling frame for some social surveys and take on the Royal Mail updates twice a year.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B2.3 Define the target population and compare it with the study population. *(Accuracy and Reliability)* ■	This comparison should be made for the population as a whole and for significant sub-populations. It gives an indication of potential coverage error, in that the study population is derived from a frame that may not perfectly enumerate the population.	Target population: adults in employment Study population: businesses on the Inter-Departmental Business Register (IDBR). A survey of a sample of businesses on the IDBR is used to estimate total employment by sector of activity. The IDBR does not include many small businesses (neither Value Added Tax (VAT) registered nor on Pay As You Earn (PAYE) system) and hence misses most self-employed – undercoverage. On the other hand, a person can be employed by more than one business – overcoverage. Also, some dead businesses may remain on the IDBR and some births may not have been added.
B2.4 Coverage error. *(Accuracy and Reliability)* ■	Coverage error arises from failure to cover adequately all members of the population being studied. This information is useful for assessing whether the coverage is sufficient for the required precision of the estimates. The population that the data covers should be included along with all known coverage problems. There could be overcoverage (where duplicate records are included) or undercoverage (where certain records are missed). Special studies can sometimes be carried out to assess the impact of undercoverage and overcoverage. These should be reported where available. If appropriate and available, quantitative measures can be used to highlight the extent of the following coverage issues: • estimated rate of undercoverage; • estimated rate of ineligible units; • estimated rate of duplicate units; and • estimated rate of misclassified units. Further information on possible coverage error is gained when the study population and target population are stratified according to key variables and then compared. This indicator can be estimated from follow-up surveys.	Three sources of known undercoverage impact on the overall Public Sector Employment (PSE) estimates. First, an issue with the local authority collection relates to Local Education Authority (LEA) maintained schools in England and Wales opting out of local authority payroll administration. As a result they are not included in the data collected. A second issue with the local authority collection is with schools in England and Wales adopting academy status. Such schools are classified as central government and data are collected from them individually. During the transition, there is the potential for under-counting or double counting. In addition, local authorities may continue to operate payrolls on behalf of academies, creating the potential for misclassification. There are currently around 1,000 academies across England and Wales. Third, PSE has known under-coverage as part of its survey of public bodies; public bodies with less than 20 employees are not surveyed. The level is approximately 500 people and this is not considered to affect the overall survey estimates. *Source: Public Sector Employment (PSE), Quality and Methodology Information, Office for National Statistics, 2012*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B2.5 Assess the likely impact of coverage error on key estimates. *(Accuracy and Reliability)* ■	This can indicate to users how reliable the key estimates are. The assessment of the likely impact of coverage error on key estimates should draw on information on coverage error for sub-groups where appropriate, and of estimates of rates of undercoverage; overcoverage; misclassifications; and ineligible units.	The court system in England and Wales records all divorces taking place in England and Wales regardless of usual residence of the two parties. It is possible that some people who are usually resident in England and Wales may undertake divorce proceedings in another country so are not included in the divorce statistics for England and Wales. Similarly, some people who are overseas residents may divorce in England and Wales. Although there are estimates for marriages abroad (using data from the International Passenger Survey (IPS) on intention of travel to get married), there is no such equivalent for divorces abroad. It is likely that the numbers of divorces taking place abroad are low (given this option is not open to those without a connection to another country) so the exclusion of such divorces will not impact the main uses of the data. *Source: Divorces, Quality and Methodology Information, Office for National Statistics, 2012*
B2.6 Describe methods used to deal with coverage issues. *(Accuracy and Reliability)* ■	Updating procedures, frequency and dates should be described, in addition to frame cleaning procedures. Also, edit checks, imputations or weighting procedures that are carried out because of coverage deficiencies should be described. This information indicates to users the resultant robustness of the survey sample or admin dataset as a representative sample of the target population.	Where a survey does not cover the whole business population, such as the Annual Business Survey (ABS), re-classification can lead to units moving out of the sample, but never into it. In the ABS, this effect is likely to be small, and is corrected for by adjusting the weights of the businesses which remain in the sample. *Source: Annual Business Survey (ABS), Quality and Methodology Information, Office for National Statistics, 2013*
B2.7 Describe the sample design. *(Accuracy and Reliability)* ▲	This provides the user with information relating to sample size and how the sampling units were selected. Common sampling methods include simple random sampling, clustered sampling, multi-stage sampling, stratified sampling and systematic sampling.	The Business Register Employment Survey sample currently contains around 80,000 businesses from across the Great Britain economy. The Inter-Departmental Business Register (IDBR) is used as the sampling frame from which a stratified random sample is drawn. The strata are defined by Standard Industrial Classification (SIC 2007) by country and by employment size, with all employment sizes of businesses being covered. The design is a stratified one stage clustered sample, where the stage 1 units (or clusters) are enterprises, or Reporting Units (RUs), and the elements in each cluster are local units. *Source: Business Register Employment Survey (BRES), Quality and Methodology Information, Office for National Statistics, 2013*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B2.8 For an ongoing survey, have there been any changes over time in the sample design methodology? *(Coherence and Comparability)* ▲	Sample design and estimation methods are often linked, though it is possible to change one and not the other. However, any change can affect the outputs. Examples of changes to details of the methodology include changes in the sample size, or a re-allocation of the sample across strata; or it might be that new stratum definitions or boundaries are introduced. Larger changes might include a wholesale redesign of the survey, for example moving from a longitudinal design to a cross-sectional one, or vice versa. Changes in estimation procedures include changing details, such as the choice of auxiliary variable(s) used in model-assisted design, the type of estimator (eg expansion, ratio or regression estimation), or some other aspect of the processing (for example, some outputs (eg Gross Domestic Product, GDP) are formed through the amalgamation of many different inputs, and any of the input sources might themselves be changed in some way. Particular care is needed in estimating the precision of change over the change in design if this disrupts, for example, rotation patterns.	Classification changes often necessitate a review and redesign of sample and estimation methods. Standard Industrial Classification (SIC 2007) - comparable with the European NACE Rev. 2 - was introduced to official statistics in the UK between 2008 and 2011. This necessitated the redesign of most official business surveys in terms of construction of new sampling strata, which are defined around industry definitions. Estimates for SIC (2007) industries also needed to be back cast, and time series now typically comprise a recent section based on the new design, a centre section based on the old strata but with estimates derived from re-calibrated microdata, and a historic section based on conversion of the existing aggregate estimates. Other changes often occur in surveys for practical purposes, for example, the Labour Force Survey (LFS) saw some fairly minor design changes - around sampling at multiple-occupancy addresses and sampling of persons over-75 households - in 2010; these both required commensurate changes in the calculation of design weights to ensure the survey responses were still being weighted in an appropriate way.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B2.9 Describe classifications. *(Relevance)* ■	This is an indicator that ensures users are informed of concepts and classifications used in compiling the output. The information should be sufficient to allow replication of data collection and compilation.	From August 2010 onwards, the Average Weekly Earnings (AWE) sample was designed, processed and weighted using the 2007 Standard Industrial Classification (SIC 2007), having used SIC 2003 previously. The United Kingdom Standard Industrial Classification of Economic Activity 2007 (UK SIC 2007) is the national version of the European system 'Classification of Economic Activities in the European Community' Rev: 2 (NACE Rev 2). UK SIC is a five digit hierarchical classification, identical to NACE down to and including the fourth digit 'class' level of the system. In certain instances, UK SIC 2007 provides a further breakdown of the class level by the addition of a national 5 digit subclass level.UK SIC classifies business according to their main economic activity, providing a uniform structure and framework for the collection, tabulation, presentation and analysis of data. Following the change to SIC 2007, all AWE historic time series were re-estimated on a SIC 2007 basis. The transition to SIC 2007 led to a substantial re-design of the Monthly Wages and Salaries Survey (MWSS) sample from August 2010 onwards, which may lead to increased sample error of the growth rates between August 2010 and July 2011. ONS publishes a table to help users of the series determine the nearest currently published equivalent to the series they previously used. *Source: Average Weekly Earnings (AWE), Quality and Methodology Information, Office for National Statistics, 2011*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B2.10 Provide a statement of the nationally/internationally agreed definitions and standards used. *(Coherence and Comparability)* ■	This indicates geographical comparability where the agreed definitions and standards are used.	Gross Domestic Product (GDP) estimates are produced in line with international standards, most notably ESA95 which is enforced for all European Union Member States through EU council regulation (EC) No 2223/96. ESA95 is in turn consistent with the United Nations System of National Accounts 1993 (SNA93). The SNA93 has recently been updated as SNA 2008 and in turn the ESA95 is being revised and will form ESA 2010, which will subsequently be implemented in the UK national accounts. Further information on the introduction of ESA 2010 is available on the ONS website. *Source: Gross Domestic Product (GDP), Quality and Methodology Information, Office for National Statistics, 2012*
B2.11 Provide a statement of international regulations that apply and any laws which have been repealed or abolished. *(Coherence and Comparability)* ■	The provision of this information allows users to assess whether international regulations or repealed or abolished laws will affect comparability.	The Products of the European Community (PRODCOM) Survey is a European initiative and is governed by EU regulation. It is a compulsory survey for UK businesses and is collected under the statutory powers of the Statistics of Trade Act 1947 for Great Britain, and Employment (NI) Order 1988 for Northern Ireland. PRODCOM is a Harmonised System (the Harmonised Commodity Description and Coding System, is a worldwide reference for classifications of external trade statistics and for customs tariffs) for the collection and publication of product statistics. *Source: Products of the European Community (PRODCOM), Quality and Methodology Information, Office for National Statistics, 2012*
B2.12 Describe any deviations from nationally/internationally agreed definitions and standards. *(Coherence and Comparability)* ■	This should include reasons for any deviations. Where there are deviations from national or international definitions and standards, these may make data less comparable with other data that conform to these agreed definitions and standards. This indicator allows users to judge whether the data are comparable to other data from the same or other geographical areas.	Although Public Sector Finances (PSF) are compiled on the bases of the National Accounts framework, some of the measures used in the bulletin are UK specific measures which have been defined by the UK Government for the purposes of fiscal policy. The methods, classification and terminology used in PSF are detailed in the following publications: • Background notes to the PSF statistical bulletin; • Background notes to the Government Deficit and Debt under the Maastricht Treaty Statistical bulletin; and • Monthly Statistics on Public Sector Finances - A Methodological Guide. *Source: Public Sector Finances (PSF), Quality and Methodology Information, Office for National Statistics, 2012*

Beyond the five dimensions of statistical output quality outlined in Table A1, the European Statistical System Handbook for Quality Reports also identifies other considerations which are relevant to measuring and reporting on quality. Further information is included on page 9.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

B3. Collection

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B3.1 For outputs that require the use of administrative sources, describe the timeliness and punctuality of the data received. *(Timeliness and Punctuality)*	Provide information on how soon after their collection the statistical institution receives the administrative data. The effects of any lack of timeliness on the statistical product should be described. Lack of punctuality in the delivery of the administrative data source should also be described. Any reasons for the delay between the scheduled and actual delivery dates of the data should be documented.	Cancer incidence and mortality in the United Kingdom brings together data from England, Wales, Scotland and Northern Ireland. Cancer incidence data for England are compiled by ONS using data supplied by the eight English cancer registries; data for Wales are collected by the Welsh Cancer Intelligence Surveillance Unit; data for Scotland are collected by the Information Services Division (ISD) Scotland cancer information programme and data for Northern Ireland are collected by the Northern Ireland cancer network. Mortality data are collected by ONS for both England and Wales, while the General Register Office (GRO) for Scotland and the Northern Ireland Statistics and Research Agency (NISRA) collect data for Scotland and Northern Ireland respectively. Previously, the cancer registries in England were obliged by the Department of Health to provide data to ONS within 18 months of the end of the registration year (January to December). For 2009, registries provided the data within 15 months and for 2010 this was reduced to 12 months to further improve the timeliness of the publications. After registrations have been received by ONS, they are processed and cleaned using validity and consistency checks (see the Accuracy section for more information). *Source: Cancer incidence and mortality in the United Kingdom, Quality and Methodology Information, Office for National Statistics, 2013*
B3.2 When using more than one source, describe the common identifiers of population units in the data and provide details of identifiers if there is none in common. *(Coherence and Comparability)*	Different sources often have different population unit identifiers. The statistical producer can utilise this information to match records from two or more sources. Where there is a common identifier matching is generally more successful.	The main administrative sources investigated by ONS for use with business statistics are Value Added Tax (VAT) and Company Accounts. In both cases, the unique identifiers in the data sets do not match with the statistical units used for official statistics. This is because the reporting structure that businesses use for reporting VAT and Company Accounts information is not necessarily the same as that required for ONS outputs. *Source: Lewis D and Woods J (2013) Issues to consider when turning to the use of administrative data: the UK experience, NTTS conference 2013, Brussels, Belgium*

= relevant to survey data based outputs; = relevant to administrative data based outputs; = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B3.3 Describe the record matching methods and processes used on the data sources. *(Accuracy and Reliability)* ■	Record matching is when different sources are brought together and units on the separate sources are matched, where appropriate, as being the same units. There are many different techniques for carrying out this process. A description of the technique (eg automatic or clerical matching) should be provided along with a description (qualitative or quantitative) of its effectiveness.	For the 2011 Census, census records were matched to the 2011 Census Coverage Survey (CCS) using a combination of exact matching, score based matching, clerical matching and clerical searching. A summary of the clerical matching methods is provided below; • Clerical matching – the manual review of pairs of records that are classified as potential matches based on their overall agreement scores. These records have scored lower than the auto-match threshold and a trained matcher is required to make a decision based on the evidence available as to whether or not the two records should be linked. • Clerical searching – individually taking the 'residuals' (unmatched records) on one of the datasets and querying the database of the second dataset for a matching record. Where potential match pairs are identified, a clerical decision is made by the matcher as to whether or not to link the two records. In the 2011 Census to CCS matching, the role of clerical matching and clerical searching was very important for the optimisation of matching accuracy. By using these methods, it ensured that the number of false positives and false negatives were kept to a minimum, thereby enabling accurate coverage adjustments to be made in the estimation process. *Source: Beyond 2011: Matching Anonymous Data, Office for National Statistics, 2013.*
B3.4 Percentage of common units when combining more than one source. *(Coherence and Comparability)* ■	This indicator provides information on the proportion of common units across two or more data sources. Only units relevant to the statistical output should be considered.	**Percentage of common units across two or more admin sources:** $\frac{\text{No. of relevant common units in the data sources}}{\text{No. of relevant unique units}} \times 100\%$ Note: The "unique units" in the denominator means that units should only be counted once, even if they appear in multiple sources. This indicator should be calculated separately for each variable. If the sources are designed to cover different populations and are combined to provide an overall picture, this should be explained. Further information on the calculation of this indicator is available as part of the ESSnet Admin Data deliverable 6.5, available here: http://essnet.admindata.eu/WorkPackage/ShowAllDocuments?objectId=4257

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example
B3.5 Percentage of consistent items for common variables in more than one source. *(Coherence and Comparability)* ■	This indicator provides information on consistent items for any common variables across sources (either admin or survey). Only variables directly required for the statistical output should be considered – basic information (eg business name and address) should be excluded. Variables in common on all sources can be compared for consistency. As well as comparing definitions, the values held for common units can also be compared, and judged as being consistent if they are within a given tolerance of each other. For numeric values, the tolerance may be given as a percentage, for example. An indicator for consistency should be calculated for each of the key variables and aggregated based on the contributions of the variables to the overall output.	**Percentage of consistent items for common variables:** $\frac{\text{No. of consistent items (within tolerance) for X variable}}{\text{Total no. of items required for X variable}} \times 100\%$
B3.6 Describe the tolerance and the reasons for this. *(Coherence and Comparability)* ■	Where values within a particular tolerance are considered consistent for a common variable(s) across more than one source, the tolerance should be stated, along with a brief explanation as to why this particular value was chosen.	Turnover values from Value Added Tax (VAT) and survey data were assumed to be consistent when they were within 5% of each other. The tolerance takes into account the difference in definition between the two sources.
B3.7 For survey estimates, what were the set, target and achieved sample sizes? *(Accuracy and Reliability)* ▲	The target sample size is the sample size specified under the sample design used for the survey. The set sample size is the number of sampling units drawn from the frame. The achieved sample size will typically differ from the set sample size because of non-response, non-contacts and ineligible units, and will differ from the target sample size if the predictions of these are not exactly correct.	For the Opinions and Lifestyle Survey (OPN), the issued sample size is 2010 addresses per month and the achieved sample is approximately 1,100 addresses per month. *Source: Opinions and Lifestyle Survey, Quality and Methodology Information, Office for National Statistics, 2012*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B3.8 Non-response and non-response bias. *(Accuracy and Reliability)* ■	Non-response is failing to obtain some or all of the information from a unit. Non-respondents may differ from respondents, which can lead to non-response bias. This error cannot usually be calculated exactly. However, a statement should be provided to indicate to users that the data are subject to non-response. Where possible, potential differences between responders and non-responders should be described and an estimate of non-response bias should be provided, along with any information on studies that have been carried out to determine this level of bias, and any steps that have been taken to reduce bias as a result. If available, information should also be provided on the strategy that is in place for making contact with non-responders (eg the number of attempts to obtain a response). Non-response (or missing values) also occurs with administrative data and similar information about potential bias in this data should be reported to users.	Response rates provide an indication of the accuracy of our final estimate. For the International Trade in Services (ITIS) Survey, the final response rate target is 85%. Non-response bias is a potential issue for all statistical surveys. Non-response bias occurs where the answers of respondents differ from potential answers of non-responders. The risk of non-response bias is minimised by efforts to maximise response rates and estimation techniques can attempt to correct for any bias that might be present. Despite this, it is not easy to quantify the extent to which non-response bias remains a problem. However, there is no evidence to suggest that non-response bias presents a particular issue for the ITIS surveys. *Source: International Trade in Services (ITIS), Quality and Methodology Information, Office for National Statistics, 2013*
B3.9 Unit response rate by sub-groups. *(Accuracy and Reliability)* ▲	The response rate is a measure of the proportion of sampled units that respond to a survey. This may indicate to users how significant the non-response bias is likely to be. Higher non-response means that the presence of bias may be more likely and also increases non-response variance. NB: There may be instances where non-response bias is high even with very high response rates, if there are large differences between responders and non-responders. For formula, please see Annex A.	The overall response rate for the General Lifestyle Survey (GLF) indicates how many full and partial interviews were achieved as a proportion of those eligible for the survey. In order to obtain the most conservative response rate measures, the denominator includes an estimate of the proportion of cases of unknown eligibility that would in fact be eligible for interview. In 2010, the overall response rate was 72%. *Source: General Lifestyle Survey (GLF), Technical Appendices 2010, Office for National Statistics*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B3.10 Key item response rates. *(Accuracy and Reliability)* ■	Unit non-response occurs when it is not possible to collect any information from a sample unit. Item non-response occurs where a value for the item in question (a particular variable) is missing or not obtained for the particular unit. Non-response at item level is an issue for both admin and survey data (although there are technically no 'responses' to the statistical producer when using admin data, missing items for particular units can be considered as non-response). Non-respondents may differ from respondents, which can lead to non-response bias. Higher non-response means that the presence of bias may be more likely and also increases the variance of estimates. However, there may be instances where non-response bias is high even with very high response rates, if there are large differences between responders and non-responders. Item non-response bias can be estimated from follow up studies, and where possible, an estimate of bias due to item non-response for key variables should also be provided. Note: there are some instances where key variables are derived from other variables so that the derived variables are typically affected by the item non-response in their component parts.	**Unweighted item response rate:** $\frac{\text{Number of units with a value for item}}{\text{Number of units in scope for the item}}$ **Weighted item response rate:** $\frac{\text{Total weighted quantity for item for responding units}}{\text{Total weighted quantity for the item for all units}}$ Note: The information required to calculate these quantities may be available for both respondents and non-respondents on the sampling frame.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B3.11 Estimated variance due to non-response. *(Accuracy and Reliability)* ▲	Non-response generally leads to an increase in the variance of survey estimates, since the presence of non-response implies a smaller sample and extra variation if the decision to take part is regarded as random. An estimate of the extra variance (on top of the normal sampling variance) caused by non-response gives useful additional information about the accuracy of estimates. It is important to note that if non-response is addressed through imputation, it can lead to the appearance of the variance being reduced, both because imputed values are usually less extreme than would be observed from sampling alone, particular if missing values are replaced with a central value and the achieved sample size is artificially.	An approximately unbiased variance estimator for VNR can be obtained by finding an approximately unbiased estimator for $V_q(\hat{\theta}^* \mid s)$ *[where q is the non-response mechanism; $\hat{\theta}^*$ is the estimated parameter, adjusted for non-response weighting adjustment or imputation; and s is the realised sample].* VNR is the additional variance from non-response. In practice, it is very difficult to estimate the non-response bias. However, it is possible to estimate the non-response variance by assuming that the bias is negligible. In the last decade, many methods were indeed proposed to estimate this variance, and some of these have been implemented in the System for Estimation of Variance due to Non-response and Imputation (SEVANI). The methodology used to develop SEVANI is based on the theory of two-phase sampling where we assume that the second phase of selection is non-response. (From http://www5.statcan.gc.ca/bsolc/olc-cel/olc-cel?catno=11-522-X20020016729&lang=eng).

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B3.12 Rate of proxy responses. *(Accuracy and Reliability)* ▲	Proxy response is a response made on behalf of the sampled unit by someone other than the unit. It is an indicator of accuracy as information given by a proxy may be less accurate than information given by the desired respondent.	Ideally, all adult members of the household should be present during the interview, so that the income section can be asked personally. However, where a member of the household is not present during the interview, another member of the household (such as spouse) may be able to provide documentary information about the absent person. The individual interview is then identified as a proxy interview. From 2001-02, the Expenditure and Food Survey (EFS)/Living Costs and Food Survey (LCF) has included households that contained a proxy interview. Around a fifth of LCF households contain at least one proxy interview. Additional Information on proxy interviews is available in the Family Spending publication. *Source: Living Costs and Food (LCF) Survey, Quality and Methodology Information, Office for National Statistics, 2012* **Rate of complete proxy responses:** $\frac{\text{Number of units with complete proxy response}}{\text{Total number of eligible units}}$ The rate of partial proxy responses can also be calculated and is the rate of complete interviews given partly by the desired respondent(s) and partly by proxy. **Rate of partial proxy responses:** $\frac{\text{Number of units with partial proxy response}}{\text{Total number of eligible units}}$
B3.13 Statement of sources of measurement error. *(Accuracy and Reliability)* ■	Measurement error is the error that occurs from failing to collect the true data values from respondents. Sources of measurement error are: the survey instrument; mode of data collection; respondent's information system; respondent; and interviewer. If information on measurement error for administrative data is available from the administrative source, this information should be included. Outputs should contain a definition of measurement error and a description of the main sources of the error.	Measurement error is the error that occurs from failing to collect the correct information from respondents. Sources of measurement error in a census include a poorly designed questionnaire, errors by field staff or errors made by the respondent. Not all of these errors can be measured. *Source: 2011 Census Statistics for England and Wales: March 2011, Quality and Methodology Information, Office for National Statistics, 2013*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B3.14 Estimate of measurement error. *(Accuracy and Reliability)* ▲	Sources of measurement error are: the survey instrument; mode of data collection; respondent's information system; respondent; and interviewer. It may be possible to estimate the size of some contributions to the measurement error through special studies, possibly as part of a training exercise, but these can be expensive and are rare. Measurement error may be detected during editing by comparing responses to different questions for the same value, for example, age and date of birth. Alternatively, other records may be consulted to detect measurement error, for example, administrative records.	In the summer of 2011, ONS carried out the Census Quality Survey (CQS) which was a small-sample voluntary survey to evaluate the information people provided on their census questionnaires. A team of interviewers visited selected households to ask census questions again in a face-to-face interview. The aim was to assess people's understanding of the questions and measure the accuracy of information collected in the census for all household members. Agreement rates for some key questions are available on the Census Quality Survey web pages. More results from the CQS will be published with future releases of 2011 Census data. *Source: 2011 Census Statistics for England and Wales: March 2011, Quality and Methodology Information, Office for National Statistics, 2013*
B3.15 Describe processes employed to reduce measurement error. *(Accuracy and Reliability)* ▲	Describe processes to reduce measurement error. These processes may include questionnaire development, pilot studies, cognitive testing, interviewer training, etc.	A number of steps were taken to maximise response rates and reduce bias and errors in the census, including: • the census questionnaire was well designed and extensively tested; • help was provided to the public via a 2011 Census website and telephone help-line; • questionnaires submitted online were automatically validated; • data captured was checked for validity and cleaned; and • edit and imputation techniques were used to estimate missing data and correct for inconsistencies. *Source: 2011 Census Statistics for England and Wales: March 2011, Quality and Methodology Information, Office for National Statistics, 2013*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B3.16 Assess differences due to different modes of collection. *(Accuracy and Reliability)*▲	A mode effect is a systematic difference that is attributable to the mode of data collection. Mode effects can be investigated using experimental designs where sample units are randomly assigned into two or more groups. Each group is surveyed using a different data collection mode. All other survey design features need to be controlled, including differential patterns of response for different modes. Differences in the survey outcomes for the different groups can be compared and assessed.	Mode error is minimised by using appropriate methods to collect data on sensitive questions - it is hoped that response bias on Labour Force Survey (LFS) earnings questions can be evaluated and perhaps reduced by asking respondents to check documentary evidence such as a payslip and recording whether such evidence was provided. There are also small but noticeable differences in the information collected by face-to-face interviewers and by telephone interviewers. Although some of the difference can be explained by respondents getting used to the interviewing process with each successive quarter's questioning, some of the difference is also due to the mode effect and it is difficult to disentangle the two causes. Estimates of employment are about 1% lower, on average, in first interviews (face-to-face) than in subsequent interviews (telephone). However, as the survey design has not changed in recent years the estimates are consistent over time, and therefore estimates of change are unaffected by these effects. *Source: Labour Force Survey (LFS), User Guide, Volume 1 – LFS Background and Methodology, 2011*
B3.17 Provide an indication of any feedback received on difficulty with answering individual questions. *(Accuracy and Reliability)*▲	Where respondents indicate that they have difficulty answering a question, this suggests that their responses to the question may be prone to measurement error. Information on difficulty in answering questions can be obtained in a variety of ways, eg from write-in answers on self-completion questionnaires, or from cognitive testing based on a purposive sample of the respondents. Any steps that have been taken to address difficulty in answering questions should be highlighted.	A redesign of the catering questionnaires was completed on behalf of the Annual Business Survey. Several areas were noted as being difficult for respondents to complete. These questions were investigated using cognitive testing techniques which resulted in some of the questions being redesigned. *Source: Annual Business Survey (ABS) Questionnaire Review – A qualitative study, Office for National Statistics, 2011*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B3.18 Are there any sensitive data items collected for which the responses could be unreliable? *(Accuracy and Reliability)* ■	It is possible that data may be influenced by the personal or sensitive nature of the information sought. If so, an assessment of the direction and amount of bias for these items should be made.	It is likely that the General Lifestyle Survey (GLF) underestimates cigarette consumption and, perhaps to a lesser extent, prevalence (the proportion of people who smoke). For example, evidence suggests (Kozlowski, 1986) that when respondents are asked how many cigarettes they smoke each day, there is a tendency to round the figure down to the nearest multiple of 10. Underestimates of consumption are likely to occur in all age groups. Under-reporting of prevalence, however, is most likely to occur among young people. To protect their privacy, particularly when they are being interviewed in their parents' home, young people aged 16 and 17 complete the smoking and drinking sections of the questionnaire themselves, so that neither the questions nor their responses are heard by anyone else who may be present. *Source: General Lifestyle Survey (GLF) Overview Report 2010, Office for National Statistics, 2012*

Beyond the five dimensions of statistical output quality outlined in Table A1, the European Statistical System Handbook for Quality Reports also identifies other considerations which are relevant to measuring and reporting on quality. Further information is included on page 9.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

B4. Processing

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B4.1 Processing error. *(Accuracy and Reliability)* ■	Processing error is the error that occurs when processing data. It includes errors in data capture, coding, editing and tabulation of the data as well as in the assignment of weights. Processing error is made up of bias and variance. Processing bias results from systematic error introduced by processing systems, for example, an error in programming coding software that leads to the wrong code being consistently applied to a particular class. Processing variance consists of random errors introduced by processing systems, which, across replications, would cancel each other out, for example random keying errors in entering data. Outputs should be accompanied by a definition of processing error and a description of the main sources of the error. Where possible, an estimate of the processing error should also be provided.	Processing error can be introduced by processes applied to the data before the final estimates are produced. It includes errors in geographical assignment, data capture, coding, data load, and editing of the data as well as in the Coverage Assessment and Adjustment (CAA) process. It is not possible to calculate processing error exactly; however, various measures were taken during each process which can be used as estimates of processing quality. *Source: 2011 Census estimates for England and Wales: March 2011, Quality and Methodology Information, Office for National Statistics, 2013*
B4.2 Describe processing systems and quality control. *(Accuracy and Reliability)* ■	This informs users of the mechanisms in place to minimise processing error by ensuring accurate data capture and processing.	Data received are reviewed using pre-determined validation gates and a process of selective editing. Data failing validation gates are prioritised by the potential impact on final product estimates. If the impact of the failure is deemed to be above a set threshold, the returned data are checked and then edited if necessary. This approach minimises both the burden on respondents and also the resource required to validate returned questionnaires. *Source: Products of the European Community, Quality and Methodology Information, Office for National Statistics, 2012*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B4.3 Scanning and keying error rates (where the required value cannot be rectified). *(Accuracy and Reliability)* ▲	Scanning error occurs where scanning software misinterprets a character (substitution error) or where scanning software cannot interpret a character and therefore rejects it (rejection error). Rejected characters can be collected manually and re-entered into the system. However, substituted characters may go undetected. Keying error occurs where a character is incorrectly keyed into computer software. Accurate calculation of scanning and keying error rates depends on the detection of all incorrect characters. As this is unlikely to be possible, the detected scanning and keying error provides an indicator of processing error, not an absolute measure.	**Estimated scanning error rate**: $\frac{\text{Number of scanning errors detected}}{\text{Total number of scanned entries}}$ **Estimated keying error rate**: $\frac{\text{Number of keying errors detected}}{\text{Total number of keyed entries}}$ It is recommended that keying and scanning errors be calculated by field (eg question response box), as this is useful information for quality reporting. They can also be calculated by character, and this is useful for management information purposes.
B4.4 Describe the imputation method(s) in place within the statistical process. *(Accuracy and Reliability)* ■	Imputation is a method for estimating missing or unusable values. The imputation method used can determine how accurate the imputed value is. Information should be provided on why the particular method(s) was chosen and when it was last reviewed.	Imputation techniques are used to estimate the value of the missing data due to non-response. Imputation is carried out for large businesses with an employment of 250 or more, and for businesses with low employment but high turnover. To calculate imputed values, ratio imputation is used. This uses the returned values of businesses within a similar industry and with a similar size to estimate the value of missing responses. For non-responding businesses with fewer than 250 employment, and without a high turnover, imputation is not carried out and totals are estimated using adjusted weights. *Source: Annual Business Survey (ABS), Quality and Methodology Information, Office for National Statistics, 2011*
B4.5 Item imputation rates. *(Accuracy and Reliability)* ■	Item non-response is when a record has partial non-response. Certain questions will not have been completed but other information on the unit will be available. Imputation can be used to compensate for non-response bias, in that the known characteristics of non-responders can be used to predict values for missing items (eg type of dwelling may be known for certain non-responders and information from responders in similar types of dwelling may be imputed for the missing items). The imputation rate can provide an indicator of possible non-response bias, in that the imputation strategy may not perfectly compensate for all possible differences between responders and non-responders.	**Item imputation rate:** $\frac{\text{Number of units where the item is imputed}}{\text{Number of units in scope for the item}}$

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B4.6 Unit imputation rate. *(Accuracy and Reliability)* ■	Unit non-response is when an eligible sample member fails to respond at all. Weighting is the usual method of compensation. However, imputation can be applied if auxiliary data are available or if the dataset is very large. Unit imputation can help to reduce non-response bias if auxiliary data are available.	**Unit imputation rate:** $\frac{\text{Number of units imputed}}{\text{Total number of units}}$
B4.7 Total contribution to key estimates from imputed values. *(Accuracy and Reliability)* ■	The extent to which the values of key estimates are informed by data that have been imputed. A large contribution to estimates from imputed values is likely to lead to a loss of accuracy in the estimates.	**Contribution to key estimates from imputed values:** $\frac{\text{Total weighted quantity for imputed values}}{\text{Total weighted quantity for all final values}}$
B4.8 Assess the likely impact of non-response/imputation on final estimates. *(Accuracy and Reliability)* ■	Non-response error may reduce the accuracy of the final estimate. An assessment of the likely impact of non-response/imputation on final estimates allows users to gauge how reliable the key estimates are as estimators of population values. This assessment may draw upon the indicator: 'total contribution to key estimates from imputed values' (see B4.7).	An intensive follow up of non-responders to the Monthly Inquiry into the Distribution and Services Sector showed that imputation worked well as long as there was response of at least 70% in most 5-digit industries (although a few industries required 85% response). Under these conditions, the estimated bias in turnover estimates between those using responses available at results time and those available after the follow up was no more than 2% for any industry.
B4.9 Proportion of responses requiring adjustment due to data not being available as required. *(Accuracy and Reliability)* ■	Where data from respondents' information systems do not match survey data requirements (eg for respondents using non-standard reference periods), responses are adjusted as required. In this process, measurement error may occur. The higher the proportion of responses requiring adjustment, the more likely the presence of measurement error. The proportion is defined as the number of responses requiring adjustment divided by the total number of responses. This is relevant where respondents consult information systems when completing a survey or providing information to the administrative data authority, and may be more applicable to business than to household surveys.	**Proportion of responses requiring adjustment:** $\frac{\text{Number of responses requiring adjustment}}{\text{Total number of responses}}$

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B4.10 Edit failure rate. *(Accuracy and Reliability)* ■	This is the number of items rejected by edit checks, divided by total number of items. This indicates possible measurement error (although it may also indicate data capture error). Note: it is sometimes useful to calculate edit failure rates at the unit level.	**Edit failure rate:** $\frac{\text{Number of key items rejected by edit checks}}{\text{Total number of items}}$
B4.11 Editing rate (for key items). *(Accuracy and Reliability)* ■	It is often useful to calculate the number of edited items as a proportion of the number of items failing the edit checks. This is referred to as the 'hit' rate and is useful to monitor the performance of the validation processes in place for the output. Editing rates may be higher due to measurement error (eg poor question wording) or because of processing error (eg data capture error). In addition, editing may introduce processing error, if the editing method is not a good strategy to compensate for values that require editing or has been implemented incorrectly.	**Editing rate:** $\frac{\text{Number of units changed by editing for the item}}{\text{Total number of units in scope for the item}}$
B4.12 Total contribution to key estimates from edited values. *(Accuracy and Reliability)* ■	The extent to which the values of key estimates are informed by data that have been edited. This may give an indication of the effect of measurement error on key estimates.	**Total contribution from edited values:** $\frac{\text{Total weighted quantity for edited values}}{\text{Total weighted quantity for all final values}}$
B4.13 Coding error rates. *(Accuracy and Reliability)* ■	Coding error is the error that occurs due to the assignment of an incorrect code to data. Coding error rate is the number of coding errors divided by the number of coded entries. It is a source of processing error. There are various ways to assess coding errors, depending upon the coding methodology employed, eg random samples of coded data may be double checked and validation checks may be carried out to screen for impossible codes. Coding error is found in all coding systems, whether manual or automated. Not all coding errors may be detected so this is more likely to be an estimated rate.	**Coding error rate:** $\frac{\text{Number of coding errors detected}}{\text{Total number of coded entries}}$

Beyond the five dimensions of statistical output quality outlined in Table A1, the European Statistical System Handbook for Quality Reports also identifies other considerations which are relevant to measuring and reporting on quality. Further information is included on page 9.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

B5. Analysis

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.1 When were the methods used for producing this estimate last reviewed? *(Relevance)* ■	The various methodological aspects of producing the statistical output should be reviewed by someone with specialist knowledge. Therefore, information should be provided on when the last review took place, any key findings, and, if known, when the next review is likely to be.	The methodology used to produce retail sales statistics is reviewed periodically by ONS. A review of the methodology and systems used in the calculation of the Retail Sales Inquiry (RSI) was conducted in 2008. The systems were all found to be working correctly. For the methodology, recommendations were made regarding the chain-linking methodology planned for implementation on RSI in 2009. The main recommendations of this review were: • Using Consumer Prices Indexes (CPI's) for deflation rather than Retail Price Indexes (RPI's) (which is closer to the requirement to create a Laspeyres volume index for Retail Sales); • Using an annual benchmark derived from the Retail Sales Index (RSI) responses, and not one from the Annual Business Inquiry (ABI) (reduces revisions, and recognises difference in concept between RSI and ABI); • Annual updating (and consequent revision) of deflation weights for the latest ABI (again closer to the requirement for a Laspeyres volume index). These recommendations were implemented and are described in the paper 'Changes to the retail sales methodology', McLaren C (2009) Economic and Labour Market Review, vol 3, no. 6. *Source: A quick guide to the Retail Sales Index (RSI), Office for National Statistics, 2011*
B5.2 Sampling error. *(Accuracy and Reliability)* ▲	Sampling error is the difference between a population value and an estimate based on a sample, and is one of the components of total survey error. Outputs derived from sample surveys should contain a statement that the estimates are subject to sampling error and an explanation of what this means. It is not usually possible to calculate the precise sampling error for a given point estimate, as this requires knowledge of the population value. In practice, an estimated sampling error is often used as an indicator. Standard errors can (in principle) be produced for a wide range of types of estimates. Note: The terms *sampling error* (in the sense of the indicator as defined above) and *standard error* are more or less interchangeable. *Sampling variance* refers to the square of the sampling error.	As the Annual Population Survey (APS) is a sample survey, it provides estimates of population characteristics rather than exact measures. In principle, many random samples could be drawn and each would give different results, due to the fact that each sample would be made up of different people, who would give different answers to the questions asked. The spread of these results is the sampling variability. *Source: Annual Population Survey (APS), Quality and Methodology Information, Office for National Statistics, 2012*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.3 Estimated standard error for key estimates of level. *(Accuracy and Reliability)* ▲	The estimated sampling error (or estimated standard error) of an estimate is an indication of its precision. It is calculated (in its simplest form) as the positive square root of the average squared deviation of sample observations from their mean, and is an estimator of the variance of the sampling distribution of a statistic. The standard error gives users an indication of how close the sample estimate is likely to be to the population value: the larger the standard error, the less precise the estimator. Estimates of level include estimates of means and totals. The method of estimating standard error depends on the type of estimator being used; there are a number of alternative methods for any particular estimator. The coefficient of variation is a measurement of the relative variability of an estimate, sometimes called the relative standard error (RSE). Estimated coefficient of variation is defined as: $\frac{\text{Standard error estimate}}{\text{Level estimate}}$ And may be expressed as a percentage. A confidence interval can also be used to indicate sampling variability and can be described as an interval within which on repeated sampling, the true value of a population parameter would lie with known (usually high) probability.	Sampling error occurs because estimates are based on a sample rather than a census. The Vacancy Survey estimates the size of this error by estimating standard errors and coefficients of variation. For the monthly estimates of total vacancies, the estimated standard errors for the level are around 20,000, or 3% expressed as a coefficient of variation (as a percentage of the estimate of the number of vacancies). This yields an approximate 95% confidence interval of +/- 40,000. For the three-month rolling average of vacancies, the estimated standard errors are around 10,000 (1.5% expressed as a coefficient of variation), giving an approximate 95% confidence interval of +/- 20,000. For a three-month average of vacancies, for a typical industrial sector, the coefficient of variation is around ten per cent. *Source: Vacancy Survey, Quality and Methodology Information, Office for National Statistics, 2012*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.4 Estimated standard error for key estimates of change. *(Accuracy and Reliability)* ▲	Where the absolute or relative change in estimates between two time points is of interest to users, it is desirable to calculate standard errors for this change. This may need to take into account the covariance between estimates from sample coordination over time. As with the standard error for key estimates of level, the standard error for key estimates of change can be used to calculate confidence intervals of the estimates of change. Coefficients of variation are not suitable for use in summarising standard errors of changes, as the value is undefined when the change is zero and potentially very large for small changes.	The Retail Sales Index (RSI) is a monthly index published by ONS measuring the value and volume of retail sales in Great Britain. A method for estimating the standard errors of movement for indices over any fixed period has been developed by ONS and applied to a number of ONS indices over the last few years, including the Index of Production (see Wood et al 2006) and the Producer Price Index (see Morris and Green 2007). The method has now also been adapted for the RSI. An indicative standard error for the whole retail sector for 2009 is estimated as 0.4% for one–month movement and 0.7% for 12–month movement. *Source: Measuring the Accuracy of the Retail Sales Index, Office for National Statistics, 2011*
B5.5 Reference/link to documents containing detailed standard error estimates. *(Accuracy and Reliability)* ▲	The reference or link should lead to detailed standard error estimates, also providing confidence intervals or coefficients of variation, if appropriate or describing how to calculate one from the other.	Standard errors are calculated for the main International Passenger Survey (IPS) variables as well as absolute and relative confidence intervals and design effects. These are published in the Travel Trends publication (Appendix E). *Source: International Passenger Survey (IPS), Quality and Methodology Information, Office for National Statistics, 2012*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.6 Describe variance estimation method. *(Accuracy and Reliability)* ■	Describing the variance estimation method provides information on what factors have been considered when calculating the estimate. The description should include factors taken into consideration (eg imputation, measurement error, etc). The method for estimating the precision (whether given as a variance, standard error or coefficient of variation/relative standard error) of estimated survey outputs should be described, along with any practical considerations that users of the outputs or microdata (if available) should bear in mind. Methods used might include explicit calculation, as is the case for many business surveys: readily available formulas are used to calculate estimates of variances; these take into account the stratified design of the survey and the estimator used. On social surveys, the combination of complex, multi-stage sampling designs and calibrated estimation means that an explicit formula is not available and calculations that draw upon a range of information about the design are used instead to estimate precision. This information is not apparent on the micro datasets made available, meaning that users would struggle to recreate the variances estimated. Instead, design factors are estimated and provided in the technical appendices, and users can apply these to obtain approximate variance estimates from the complex design - see B5.6. Replication methods (eg the Bootstrap) are also sometimes used to estimate precision for estimators for which no algebraic formulation is available. Details of the methods, assumptions and conditions applied to produce such estimates should be supplied. In all circumstances, survey design details that affect the precision of estimates should be stated. These include the sample design and estimator applied, and how non-response is handled, as this can affect the observed precision. For the latter, if a central measure is imputed for non-responders, for example, then any measures of precision that includes these values is likely to overstate the precision, as the imputed values would never contain the outliers that would occasionally be observed as real sample responses and the data would contain more central values than would naturally be observed with full response. Information on estimating the variance for administrative data based estimates is available as part of the outcome of the ESSnet Admin data, specifically deliverable 6.3 http://essnet.admindata.eu/WorkPackage?objectId=4257	The General Lifestyle Survey (GLF) uses a multi-stage sample design, which involves both clustering and stratification. The complexity of the design means that standard errors estimated on the basis of a simple random sample design will not reflect the true variance in the survey estimates. Clustering can lead to a substantial increase in standard error, if the households or individuals within the Primary Sampling Units (PSUs) are relatively homogenous but the PSUs differ from one another. By contrast, stratification tends to reduce the standard error and is particularly effective when the stratification factor is related to the characteristics of interest on the survey. Due to the complexity of the GLF sample design, the size of the standard error depends on how the characteristic of interest is spread within and between the PSUs and strata. The method used to estimate the standard errors for the survey takes this into account. It explicitly allows for the fact that the estimated values (percentages and means) are ratios of two survey estimates: the number with the characteristic of interest is the numerator (y) and the sample size is the denominator (x), both of which are subject to random error. *Source: General Lifestyle Survey (GLF), Quality and Methodology Information, Office for National Statistics, 2012*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.7 Design effects for key estimates. *(Accuracy and Reliability)* ▲	The design effect (*deff*) indicates how well, in terms of sampling variance, the sampling method used by the survey fares in comparison to simple random sampling. Design effect = $\frac{\text{Var(estimate)}}{\text{Var}_{SRS}\text{ (estimate)}}$ where: Var(estimate) is the variance of the estimate under the complex survey design. Var_{SRS} (estimate) is the variance that would be obtained if the survey used simple random sampling. The design factor (*deft*) is the square root of the design effect. The *deft* is the ratio of the standard error under the complex design to the standard error under a simple random sample. Care should be taken to distinguish between the similar *deff* and *deft* terms. Sometimes measures of precision for complex survey designs can be estimated by the survey organisation, but would not be easy to replicate by users who have access to the microdata, as these files do not contain all the information required (eg this might include information about implicit strata created in the variance estimation process). To assist users, *deffs* or *defts* can be calculated by the survey organisation for a range of outputs (including a range of different variables and in different domains) and published as a technical annex to the main survey outputs. Users, who wish to calculate estimates in their own domains of interest from the microdata, can then estimate a variance for their outputs by first calculating a variance as though the data came from a simple random sample using the usual formulas found in survey estimation text books, and then scaling this by a design effect chosen as appropriate from those published: Var(estimate) = Design effect x Var_{SRS}(estimate) There is no need to give design effects where the sampling errors are already provided.	The design factor, or *deft*, is the ratio of the standard error of an estimate calculated taking into account the complex design relative to the standard error that would have resulted had the survey design been a simple random sample of the same size. Design factors are estimated for some of the main variables on Living Costs and Food (LCF) survey. These are published in Appendix B of Family Spending. The size of the *deft* varies between survey variables according to the degree to which a characteristic is clustered within Primary Sampling Units (PSUs), or is distributed between strata, and the impact of the weighting. For a single variable, the size of the deft also varies according to the size of the subgroup on which the estimate is based and on the distribution of the subgroup between PSUs and strata. *Defts* below 1.0 show that the standard errors associated with the complex design are lower than those associated with the simple random sample probably due to the benefits of stratification. *Defts* greater than 1.0 show the survey has produced less precise estimates than would be obtained from a comparable simple random sample because of the effects of clustering and weighting. *Source: Living Costs and Food Survey (LCF), Quality and Methodology Information, Office for National Statistics, 2012* …….. Tables of design effects or design factors can be found in Annex A of the LFS User Guide (Vol. 1) http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/labour-market-statistics/index.html

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.8 Effective Sample Size. *(Accuracy and Reliability)*▲	The Effective Sample Size (abbreviated to *neff*) indicates the impact of complex sampling methods on standard errors. For a given survey estimate, it gives the sample size that would be needed under simple random sampling to achieve the same precision as observed for the complex design. It is measured as the ratio of the achieved sample under the complex design to the design effect for each estimate and can be different estimates, variables or subgroups: Effective sample size = $\frac{n_{achieved}}{deff}$ where: $n_{achieved}$ is the achieved sample size under the complex design. *deff* is the Design Effect	The method of sampling and the consequent weighting affect the sampling errors of the survey estimates. The effect can be shown by calculating the Effective Sample Size which gives the size of a simple random sample, equivalent in precision to the complex sample design actually used. The Effective Sample Size will vary slightly from one month to another with the proportions of interviews in different sized households. On average, the Effective Sample Size of the Opinions and Lifestyle survey is 84% to 86% of the actual sample size of individuals. An achieved sample of 1,800 individual adults in the Opinions and Lifestyle survey is equivalent to simple random sample of about 1,500. *Source: http://www.ons.gov.uk/ons/about-ons/products-and-services/opn/index.html*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.9 What method of estimation was used to produce the population parameter? *(Accuracy and Reliability)* ▲	Data from sample surveys can be used to estimate unknown population values. Estimates can be calculated directly or by applying a sampling weight. The method of estimation has a bearing on the accuracy of estimates. Greater accuracy can often be achieved by using an estimator that takes into account an assumed relationship between the variables under study and auxiliary information. The method used should be explained.	As it is not feasible to survey every business in the population, it is necessary to weight the data to provide estimates for the full population. In strata that are not fully enumerated, two weights are applied: 1. 'Design' weight, also known as the 'a' weight. This takes into account the ratio of the size of a sample to the size of the population from which the sample is selected. 2. 'Calibration' weight, also known as the 'g' weight. This takes into account how large a sample's register turnover is compared to the register turnover of the population from which the sample is selected. The Monthly Business Survey (MBS) uses a mix of separate and combined ratio estimation for each specific industry. In combined ratio estimation, a common model is used for estimation for all sampled businesses in an industry irrespective of their employment strata. In separate ratio estimation, a separate model is used for each employment stratum. A third weight incorporates a Winsorisation of outliers to reduce their effect on estimates, and is applied to all data returns. *Source: Monthly Business Survey, Quality and Methodology Information, Office for National Statistics, 2012*
B5.10 Describe the estimation models used in each output. *(Accuracy and Reliability)* ■	It may be useful to describe the estimation models used and to record the assumptions underlying each model.	The standard errors for this index were calculated using a parametric bootstrap. There is an assumption that the data follow a multivariate normal distribution. The calculation of the variance-covariance matrix required to produce these estimates was simplified by making the assumption that the sample selection processes for each input to this index are independent of each other.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.11 Robustness to model misspecification. *(Accuracy and Reliability)* ■	Model misspecification error is the error that occurs when assumed or fitted models do not satisfactorily approximate the population or concept being modelled. It is not usually possible to measure model misspecification error. However, estimates produced using models should be accompanied by a description of the model and the model assumptions and an assessment of the likely effect of making these assumptions on the quality of estimates. Where possible, information should be provided on studies that have been carried out to evaluate how well model assumptions give an indication of model assumption error. Using models to improve the precision of estimates generally introduces some bias, since it is impossible to reflect completely the underlying survey populations through modelling. If available, an estimate of the bias introduced by models should also be provided.	The estimates in this report were calculated by making assumptions about the true underlying population. The model assumed for the population is the most appropriate based on the information available. However, it is impossible to completely reflect the true nature of the population through modelling. Because of this, there will be a difference between the modelled and true populations. This difference introduces error into the survey estimates.
Note: Considerations should be given to how time-series can be visualised. See Annex B.		

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.12 Analysis of variance to assess the quality of the trend estimates. *(Accuracy and Reliability)* ■	The analysis of variance (ANOVA) compares the variation in the trend component with the variation in the seasonally adjusted series. The variation of the seasonally adjusted series consists of variations of the trend and the irregular components. ANOVA indicates how much of the change of the seasonally adjusted series is attributable to changes in primarily the trend component. The statistic can take values between 0 and 1 and it can be interpreted as a percentage. For example, if ANOVA=0.716, this means that 71.6% of the movement in the seasonally adjusted series can be explained by the movement in the trend component and the remainder is attributable to the irregular component. This indicator can also be used to measure the quality of the estimated trend.	$$\text{ANOVA}=\frac{\sum_{t=2}^{n}\left(D12_t - D12_{t-1}\right)^2}{\sum_{t=2}^{n}\left(D11_t - D11_{t-1}\right)^2}$$ $D12_t$ = data point value for time t in table D12 (final trend series) of the analytical output. $D11_t$ = data point value for time t in table D11 (final seasonally adjusted series) of the output. Notes: 1. If annual totals of the non-seasonally adjusted and seasonally adjusted series are constrained to be equal, the D11A table is produced and D11A point values are used in place of D11 in the above equation. The tables mentioned above (D12, D11, D11A and A1) are automatically generated by X-12-ARIMA and X-13ARIMA-SEATS.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.13 Months (or quarters) for cyclical dominance. *(Coherence and Comparability)* ■	The months for cyclical dominance (MCD) or quarters for cyclical dominance (QCD) are measures of volatility of a monthly or quarterly series respectively. This statistic measures the number of periods (months or quarters), that need to be spanned for the average absolute percentage change in the trend component of the series, to be greater than the average absolute percentage change in the irregular component. For example, an MCD of 3 implies that the change in the trend component is greater than the change in the irregular component for a span at least three months' long. The MCD (or QCD) can be used to decide the best indicator of the long-term change in the seasonally adjusted series – if the MCD=3, the three-months-on-three-months growth rate will be a more stable than the month-on-month growth rate. The price of the stability is reduced timeliness. The lower the MCD (or QCD), the less volatile the seasonally adjusted series is and the more appropriate the month-on-month growth rate is as a measure of long-term change. The MCD (or QCD) value is automatically calculated by X-12-ARIMA and X-13ARIMA-SEATS and is reported in table F2E of the analytical output. For monthly data the MCD takes values between 1 and 12, for quarterly data the QCD takes values between 1 and 4.	MCD (or QCD) = d for the first value of d after which all — for where for multiplicative series $\bar{C}_d = \frac{1}{n-1}\sum_{t=2}^{n}\left\lvert\frac{C_t}{C_{t-d}} - 1\right\rvert$ $\bar{I}_d = \frac{1}{n-1}\sum_{t=2}^{n}\left\lvert\frac{I_t}{I_{t-d}} - 1\right\rvert$ and for additive series $\bar{C}_d = \frac{1}{n-1}\sum_{t=2}^{n}\lvert C_t - C_{t-d}\rvert$ $\bar{I}_d = \frac{1}{n-1}\sum_{t=2}^{n}\lvert I_t - I_{t-d}\rvert$ where I_t is the final irregular component from table D13 of the X-12-ARIMA or X-13ARIMA-SEATS output and C_t is the final trend component from table D12.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.14 The M7 statistic. *(Accuracy and Reliability)* ■	This is a quick indicator of whether the original series has a seasonal pattern or not. Low values indicate clear and stable seasonality. The closer to 0, the better the seasonal adjustment. Values around 1 imply that the series is marginally seasonal. An M7 value above 1 indicates that the series is non-seasonal. The M7 statistic is calculated by X-12-ARIMA and X-13ARIMA-SEATS and can be obtained from the F3 table of the analytical output. Note: this measure cannot be used in isolation, as it may be affected by other effects in the series such as outliers that have not been appropriately dealt with. It should therefore be used alongside all of the other diagnostics in the X-12-ARIMA or X-13ARIMA-SEATS output. Note: also that this measure is only calculated if seasonal adjustment is performed with the X-11 algorithm. It is not produced if seasonal adjustment if done using SEATS.	$$M7=\sqrt{\frac{1}{2}\left(\frac{7}{F_s}+\frac{3F_m}{F_s}\right)}$$ Where: $$F_M=\frac{S_B^2/(N-1)}{S_R^2/(N-1)(k-1)} \qquad F_S=\frac{S_A^2/(k-1)}{S_R^2/(n-k)}$$ F_M = F-statistic capturing moving seasonality. S_B^2 is the inter-year sum of squares, S_R^2 is the residual sum of squares, N is the number of observations. F_S = F-statistic capturing stable seasonality. S_A^2 is the variance due to the seasonality factor and S_R^2 is the residual variance. k = number of periods within a year, k = 4 for quarterly series, k = 12 for monthly series. M7 compares on the basis of F-test statistics the relative contribution of stable (statistic F_M) and moving (statistic F_S) seasonality.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.15 Comparison of annual totals before and after seasonal adjustment. *(Accuracy and Reliability)* ■	This is a measure of the quality of the seasonal adjustment and of the distortion to the seasonally adjusted series brought about by constraining the seasonally adjusted annual totals to the annual totals of the non-seasonally adjusted series. It is particularly useful to judge if it is appropriate for the seasonally adjusted series to be constrained to the annual totals of the non-seasonally adjusted series.	Formula (multiplicative models): $$\frac{1}{n}\sum_{i=1}^{n}\left\lvert (E4^{D11}_{total(t)} - 100)\right\rvert$$ $E4^{D11}_{total(t)}$ = unmodified ratio of annual totals for time *t* in table E4 of X-12 Arima. E4 is the output table that calculates the ratios of the annual totals of the non-seasonally adjusted series to the annual totals of the seasonally adjusted series for all the n years in the series. Formula (additive models): $$\frac{1}{n}\sum_{i=1}^{n}\left\lvert \frac{E4^{D11}_{t}}{D11_{total(t)}}\right\rvert$$ $E4^{D11}_{total(t)}$ = unmodified difference of annual totals for time *t* in table E4. For additive models E4 is the output table that calculates the difference between non-seasonally adjusted annual total and seasonally adjusted annual totals for all the *n* years in the series. The tables above (tables E4 and D11) are automatically generated by X-12-ARIMA and X-13ARIMA-SEATS.
B5.16 Normality test. *(Accuracy and Reliability)* ■	The assumption of normality is used to calculate the confidence intervals from the standard errors derived from the regARIMA model residuals. Consequently, if this assumption is rejected, the estimated confidence intervals will be distorted even if the standard forecast errors are reliable. A significant value of one of these statistics (Geary's and Kurtosis) indicates that the standardised residuals do not follow a standard normal distribution, hence the reported confidence intervals might not represent the actual ones. (X-12-ARIMA and X-13ARIMA-SEATS tests for significance at the 1% level).	$$\alpha = \frac{\frac{1}{n}\sum_{i=1}^{n}\left\lvert X_i - \overline{X}\right\rvert}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}}$$ $$b_2 = \frac{m_4}{m_2^2} = \frac{n\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^4}{\left(\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right)^2}$$ The above two statistics (Geary's and Sample Kurtosis respectively) test the regARIMA model residuals for deviations from normality. M_4 = fourth moment. M_2^2 = square of the second moment.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

<table>
<tr><th>Quality Measure / Indicator and Quality Theme</th><th>Description</th><th>Example / Formula</th></tr>
<tr><td>B5.17 P-values.

(Accuracy and Reliability)</td><td>This is found in the 'Diagnostic Checking Sample Autocorrelations of the Residuals' output table generated by X-12-ARIMA. The p-values show how good the fitted model is. They measure the probability of the Autocorrelation Function occurring under the hypothesis that the model has accounted for all serial correlation in the series up to the lag. P-values greater than 0.05, up to lag 12 for quarterly data and lag 24 for monthly data, indicate that there is no significant residual autocorrelation and that, therefore, the model is adequate for forecasting.</td><td>
DIAGNOSTIC CHECKING
Sample of Autocorrelation Residuals
<table>
<tr><th>Lag</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th><th>11</th><th>12</th></tr>
<tr><td>ACF</td><td>-0.06</td><td>0.12</td><td>-0.1</td><td>-0.31</td><td>-0.25</td><td>-0.13</td><td>0.21</td><td>-0.09</td><td>0.21</td><td>-0.05</td><td>0.09</td><td>-0.15</td></tr>
<tr><td>SE</td><td>0.19</td><td>0.19</td><td>0.2</td><td>0.21</td><td>0.22</td><td>0.23</td><td>0.24</td><td>0.24</td><td>0.24</td><td>0.24</td><td>0.24</td><td>0.24</td></tr>
<tr><td>Q</td><td>0.09</td><td>0.56</td><td>0.86</td><td>4.04</td><td>6.2</td><td>6.8</td><td>8.51</td><td>8.86</td><td>10.8</td><td>10.89</td><td>11.31</td><td>12.44</td></tr>
<tr><td>DF</td><td>0</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td></tr>
<tr><td>P</td><td>0</td><td>0</td><td>0.355</td><td>0.133</td><td>0.102</td><td>0.147</td><td>0.13</td><td>0.181</td><td>0.148</td><td>0.208</td><td>0.255</td><td>0.256</td></tr>
</table>
Model fitted is: ARIMA (0 1 1)(0 1 1)
(Quarterly data so 12 lags will be extracted)
ACF = AutoCorrelation Function.
SE = standard error.
Q = Liung-box Q statistic.
DF = degrees of freedom.
P = p-value of the Q statistic.
The model will not be adequate if any of the p-values of the 12 lags is less than 0.05 and will be adequate if all the p-values are greater than 0.05.
Because we are estimating two parameters in this model (one autoregressive and one moving average parameter), no p-values will be calculated for the first two lags as two DFs will be automatically lost hence we will start at lag 3. Here all the p-values are greater than 0.05, eg at lag 12 Q=12.44, DF=10 and p-value = 0.256, therefore this model is adequate for forecasting.
</td></tr>
<tr><td>B5.18 Percentage forecast standard error.

(Accuracy and Reliability)</td><td>The percentage forecast standard error is an indication of the efficiency of an estimator. It is calculated as the ratio of the square root of the mean square error of the estimator to the predicted value. The larger the forecast standard error, the wider the confidence interval will be.</td><td>$$\frac{\text{forecast standard error}}{\text{forecast}} \text{x}100$$</td></tr>
<tr><td>B5.19 Percentage difference of constrained to unconstrained values.

(Accuracy and Reliability)</td><td>This indicates the percentage difference between the constrained and the unconstrained value. This needs to be calculated for each data point being constrained. The average, minimum and maximum percentage differences (among all data points of all the series) of the unconstrained should be calculated to give an overview of the effect of constraining.</td><td>$$\left(\frac{\text{constrained data point}}{\text{unconstrained data point}} - 1\right) \text{x } 100$$</td></tr>
</table>

= relevant to survey data based outputs; = relevant to administrative data based outputs; = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.20 All discontinuities are flagged and explained. *(Coherence and Comparability)* ■	The comparability of data over time may be compromised by discontinuities. Identification of a discontinuity is a judgement based on evidence that may be statistical, contextual or both.	From 1997 onwards, note that for 63.3 (activities of travel agencies and tour operators; tourist assistance activities not elsewhere classified) total turnover, total purchases and approximate gross value added at basic prices now include monies received and paid in respect of tour operators' invoices. From 2000, the coverage of the survey has been extended to include Divisions 02 and 05, Forestry and Fishing. A level shift has been incorporated in the seasonal adjustment model so that the seasonal pattern is properly estimated, but the level shift is left in the seasonally adjusted series.
B5.21 Describe and assess the impact on data of any real-world events or changes in methods over time. *(Coherence and Comparability)* ■	This is a description of changes in concepts, definitions, classifications and methods and their causes, whether due to statistical considerations (for example, a change in measurement instrument or in the way that it is applied) or real-world events. The description should include an assessment of the impact of discontinuities on data comparability over time.	These changes include industrial action taken by registration officers, which affected the quality of information about deaths from injury and poisoning. This action meant that details normally supplied by Coroners were not available and the statistics were significantly affected. Information normally requested to aid coding of underlying cause of death was also unavailable. Figures on injury and poisoning for 1981, with the exception of suicides, must therefore be treated with caution.
B5.22 Describe time series available. *(Coherence and Comparability)* ■	Description of which data from previous periods are available to users.	The time series contain data on the following: public sector finance, central government revenue and expenditure, money supply and credit, banks and building societies, interest and exchange rates, financial accounts, capital issues, balance sheets and the balance of payments. Monthly data are available for this series going back to April 2002.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Note: The following quality measures and indicators (B5.23 to B5.29) are only relevant to those outputs that require index number construction in their production.		
Quality Measure / Indicator and Quality Theme	**Description**	**Example / Formula**
B5.23 Provide information on the type of index produced and the formula used to produce it. *(Relevance)* ■	This indicator should comment on what quantity the index is intended to measure and the type of index, for example, common types of index are a price index and a volume index. Also, the formula used to produce the index should be included (eg Laspeyres, Paasche, Fisher, Lowe). The choice of index formula can have a significant effect on the index series produced and as such details should be provided as to why this method was chosen (eg in line with international best practice).	The Consumer Prices Index (CPI) is a measure of the change in the price level of consumer goods and services bought for the purpose of consumption by households in the UK and foreign visitors to the UK. Consumer price inflation is defined as the rate of change in the general level of prices of goods and services. Within each year, the CPI is a fixed weight price index: it measures the average change in the prices of goods and services in a basket of fixed composition, quantity and as far as is possible quality. This is often summarised by saying that the CPI uses a *fixed basket*. The index $I_{0,t}$ at time period t based on time period 0 is a Laspeyres-type or fixed base weight index, being the weighted sum of price relatives of the basket commodities at a given time period as a percentage of its price on the base period: $$I_{0,t} = 100 \times \sum_i \left(\frac{p_{ti}}{p_{0i}} \right) . w_{i,b}$$ where: P_{it} = price for i^{th} item at time period t P_{io} = price for i^{th} item time period 0 w_{ib} = weight of ith item purchased in the weight reference period b In principle, the sum should be taken over every possible good or service that is within the scope of the CPI, and the price measured in every outlet that supplies each good or service. In practice, only a sample of prices can be collected. *Source: Consumer Price Indices Technical Manual 2012 Edition, Office for National Statistics, 2012*
B5.24 Describe the methods and processes used in index number construction. *(Relevance)* ■	A general description of the methods and processes used should be provided. More technical documentation and reports can be linked where necessary.	The Consumer Prices Index (CPI) and Retail Prices Index (RPI) are annually chain-linked indices: Each year, the basket is reviewed; some items are added and some taken out according to consumer trends; weights are also updated. Each year's indices are then linked together to produce a continuous index series covering many years. More information can be found in the ONS CPI Technical Manual. *Source: Inflation – Consumer Prices Index (CPI) and Retail Prices Index (RPI), Quality and Methodology Information, Office for National Statistics, 2011*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.25 How often is the base period updated and why is the index series re-based this frequently? *(Relevance)* ■	It is considered good practice to re-base an index series at least every five years. If the frequency of updating the base period is not thought to be ideal, this should be highlighted to users, alongside reasons for not being able to re-base the index series more often. The method for determining the base should be clearly stated, for example that the 100 may be set on the annual average of monthly or quarterly estimates within the basis year.	Every five years, Producer Price Indices (PPIs) are rebased, and their weights updated to reflect changes in the industry. The following article Rebasing the Producer Price Index (PPI) (including trade prices) and Services Producer Price Index (SPPI) onto 2010=100 informs users about work underway to rebase PPIs from a 2005=100 basis to a 2010=100 basis and update the weights. PPIs will move to a 2010=100 basis from autumn 2013. More information about the impact of rebasing will be published as the project progresses and will be drawn to users' attention in the regular statistical bulletin. *Source: Producer Price Index (PPI), Office for National Statistics, 2013*
B5.26 If weights are used within the construction of the index series, comment on how often the weights are updated and why they are updated this frequently. *(Relevance)* ■	Weights are an important element of an index series and efforts should be made to obtain the best possible weighting information. It is important to consider whether the weights used are fit for purpose, including the reliability of the source of the weights. The quality of the weights should be re-examined periodically to see whether there are better sources of weighting information. Possible improvements should be highlighted, including information on and/or links to any ongoing work.	Each year ONS reviews the basket of goods and services for the Consumer Price Index (CPI) and Retail Price Index (RPI) and updates the weights used to aggregate price changes. Annual updates to the weights are necessary to ensure that the CPI and RPI remain representative of current household expenditure patterns. In addition, weights also need to be updated to reflect the introduction of new items and the removal of old ones. The updating of the weights coincides with the annual review of the CPI and RPI basket of goods and services. The 2012 review of the CPI and RPI baskets is described in, 'Consumer Prices Index and Retail Prices Index: The 2012 Basket of Goods and Services'. *Source: Consumer Prices Index (CPI) and Retail Prices Index (RPI) - Updating Weights for 2012, Office for National Statistics, 2012*
B5.27 If deflators are used, how often are they reviewed and why are they reviewed this frequently? *(Relevance)* ■	If used, the deflators should be reviewed regularly; preferably every year, but not less than every five years. If the frequency of reviewing deflators is not thought to be ideal, this should be highlighted to users, alongside reasons for not being able to carry out reviews more often (eg lack of resource). Deflators should match the domain of the series as closely as possible. Information should be provided on whether the deflator is specific to the purpose for which it is used, or whether it is generic.	In 2011, the use of deflators in the National Accounts was subject to a review. Where goods and services are provided to and consumed by households, the price index needed to deflate the current price series was switched from the Retail Prices Index (RPI) to the Consumer Prices Index (CPI). The switch to the CPI presents a number of improvements; these include • The coverage of the CPI matches the National Accounts more closely than the coverage of the RPI • The weights used in the CPI are taken from the National Accounts; the weights for the RPI are taken from an expenditure survey • The CPI is produced to comparable international standards; the RPI is unique to the UK Deflators are reviewed on a regular basis with a prioritisation process to assess those in most need of review. Reviews are carried out within the five year recommendation period.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.28 Standard errors of index number and how the accuracy of the index series is assessed? *(Accuracy and Reliability)* ■	It is considered good practice to produce a regular estimate of the accuracy of an index series. This may be calculated as a measure of accuracy for the level of the index, or the growth or both. However, it is recognised that this may be difficult to achieve. Therefore the measurement of the accuracy of the index series should be commented on, and links to relevant methodological papers should be provided where necessary. If standard errors can be produced these should be provided (for more information on calculating the estimated standard error for key estimates of change, see indicator *B5.3*).	The standard error of an index movement is a measure of the spread of possible estimates of that movement likely to be obtained when taking a range of different samples of retail companies of the same size. This provides a means of assessing the accuracy of the estimate: the lower the standard error, the more confident one can be that the estimate is close to the true value for the retail population. An approximate 95% confidence interval for the index movement is roughly twice the standard error. The paper 'Measuring the accuracy of the Retail Sales Index,' written by Winton, J and Ralph, J (2011), reports on the calculation of standard errors for month-on-month and year-on-year growth rates in the RSI as well as providing an overview of standard errors and how they can be interpreted. • The standard error for year-on-year growth in all retail sales volumes is 0.7%. This means that the year-on-year growth rate for all retail sales volumes falls within the range 0.3 ± 1.4% on 95% of occasions. • The standard error for month-on-month growth in all retail sales volumes is 0.4%. This means that the month-on-month growth rate for all retail sales volumes falls within the confidence interval -0.1 ± 0.8 on 95% of occasions. *Source: Retail Sales, Statistical Bulletin, Office for National Statistics, 2013*
B5.29 How does the index compare with similar products? *(Coherence and Comparability)* ■	Where possible, index series should be reviewed against comparable measures (for example similar indices or related outputs produced either within the same department or by other organisations) and similarities and differences should be highlighted. The extent to which any differences have been investigated should be described and any relevant links should be provided.	There can sometimes be differences in growth rates between the Retail Sales Inquiry (RSI) and the retail component of the Index of Services (IoS). This is due to the fact that the RSI measures the retail sales of Great Britain whereas the IoS measure change in Gross Value Added (GVA) for the UK retail industry. For example, because there is usually a larger mark up by retailers on non-food items than on food items, non-food retailers get a larger weight in the IoS (where the mark up causes an increase to GVA) than in the RSI. The retail component of the IoS does use RSI data in its compilation. To produce the UK GVA data is required by the IoS from the RSI. The data is benchmarked and aggregated using GVA weights (rather than the turnover weights used by the RSI) to produce an estimate for the change in gross value added for total retail. *Source: Index of Services Summary Quality Report, Office for National Statistics, 2009*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.30 Describe any confidentiality agreements or assurances given to protect participants' / respondents' confidentiality. *(Accessibility and Clarity)* ▲	This can be a broad textual description, or for fuller information it may display the actual agreement or assurance given to participants. The description may refer to standard agreements, such as the Statistics of Trade Act 1947, which govern the confidentiality of individual data.	Respondents to the Labour Force Survey (LFS) are given a confidentiality pledge. This provides an assurance that their data will be used in strict confidence to produce anonymous statistics for decision making in government and research purposes only. No data that could identify them or their household will be passed to other civil servants or local authorities, commercial organisations or the press. The Data Protection Act 1998, the Statistics and Registration Service Act 2007 and the Code of Practice for Official Statistics require ONS to protect the confidentiality of all personal and private data provided for statistical purposes.
B5.31 Describe how data collection security was ensured. *(Accessibility and Clarity)* ▲	This should be a broad description of how data collection security was ensured (for example, through secure data capture systems or secure data transfer).	Royal Mail services were used to deliver the 2011 Census forms and to log their return. Returned paper questionnaires were scanned by the Royal Mail to register their unique ID allowing them to be tracked. Processes were in place to ensure shipments of forms were tracked. Capture sites were subject to physical and procedural audits. Online data collection systems were sited in specific locations and building entry was controlled, (CCTV was in place). Encryption was applied to the online internet forms. Field workers used laptops with encrypted hard disks, which were securely wiped after the exercise. Confidentiality awareness training was provided and each individual recruited for the purpose was required to sign a formal Census Confidentiality Undertaking. *Source: 2011 Census Security: Report of the Independent Review Team, Office for National Statistics, 2011*
B5.32 Describe how processing security was ensured. *(Accessibility and Clarity)* ■	This should be a broad description of how processing security was ensured (for example, through secure office systems or secure access).	A digital image of each 2011 Census form is captured for processing. A microfilm version will be retained in strict confidentiality to be released after 100 years as a public record of great interest to family historians. The paper forms will be pulped and recycled. At the end of the data processing exercise, it is planned that all IT equipment used in the processing of 2011 Census information will be either cleansed or destroyed, and that all building which are no longer to be used for census purposes will be securely decommissioned. *Source: 2011 Census Security: Report of the Independent Review Team, Office for National Statistics, 2011*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.33 What user needs have been considered when assessing which statistical disclosure control methods to apply? *(Relevance)* ■	Producers of statistics should consider user needs when considering what form of statistical disclosure control they should employ. User needs that should be considered may include levels of classification, geographies, etc.	As well as producing tables of turnover by industry at Standardised Industrial Classification (SIC) two-digit level, there was also a requirement to produce bespoke tables when requested. To protect against the identification of an individual business, the statistical disclosure control method applied a threshold rule at Enterprise Group level and a modified dominance rule (p% rule) at Reporting Unit level. Cells failing either of these rules were suppressed, and to avoid disclosure by differencing, additional cells were selected for secondary suppressions manually. Only a small number of cells were suppressed so the tables are still of great use and confidentiality is maintained.
B5.34 How were records anonymised, ie. removal of direct identifiers? *(Accessibility and Clarity)* ■	Examples of direct identifiers include name, address, National Insurance number, etc. The removal of identifiers does not in itself ensure data confidentiality; additional measures to protect the data may still be required.	Names, telephone numbers and addresses were recorded for back-checking purposes but this information was kept in a separate file. Where names and addresses were recorded on paper, these were separated from the responses when questionnaires were returned to the office.
B5.35 Describe in broad terms the statistical disclosure control method(s) applied. *(Accessibility and Clarity)* ■	This textual description should be broad enough to give an indication of the methods employed, without allowing the user to unpick the methods and derive any disclosive data. For tabular data, reference should be made to both pre- and post-tabulation statistical disclosure control methods employed.	MICRODATA: Person type variables potentially revealing sexual identity of those respondents not in a civil partnership were recoded in the 12-month Annual Population Survey (APS) Well-being End User Licence (EUL) dataset. In addition, the health variable was removed. Households larger than ten persons were also taken out as standard practice for EUL datasets and all income and salary variables were top-coded. TABULAR DATA: Disclosure is a particularly sensitive issue in business surveys, given the commercial confidentiality of the data collected. In general, the following rules are applied to the PRODCOM estimates, the exception being if the respondent is willing to allow the data to be published. **Threshold** – across the sampled units used to produce the estimates, there must be a certain number of enterprise groups in the fully enumerated cells or at a certain number of groups in the sampled cells, otherwise data are deemed disclosive. **Dominance rule** - estimated sum of all the reporting units in the cell (including those not included in the sample) minus the largest two reporting units included in the sample, must be at least a certain percentage of the largest reporting unit value sampled. These rules are generally common to all business surveys. *Source: Products of the European Community (PRODCOM), Quality and Methodology Information, Office for National Statistics, 2012*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.36 Has any evidence been compiled in support of disclosure protection settings? *(Accessibility and Clarity)* ■	This should be a statement on the rationale behind the disclosure controls applied. Data providers should try to avoid setting arbitrary threshold levels (eg suppressing all cells smaller than 10 without evidence that these data actually represent a disclosure risk).	In 2005, the ProLife Alliance made a Freedom of Information (FOI) request to the Department of Health for data on abortion statistics under certain categories of abnormality. These data had been suppressed due to the cells having values between 0 and 9. The High Court ruled in 2011 that the data should be released as there was insufficient evidence to suggest that these data were personal information. The Information Commissioner's Office (ICO) Code therefore recommends a 'motivated intruder test' be carried out to test whether anyone can actually be re-identified in that data in practice. *Source: Anonymisation: Managing Data Protection Risk Code of Practice, Information Commissioner's Office, 2012* Subsequent to this, ONS have carried out a series of 'intruder tests' on Census and End User Licence data in support of their disclosure control settings.
B5.37 Describe the impact of statistical disclosure control methods on the quality of the output. *(Accuracy and Reliability)* ■	This should be a broad textual statement of impact on quality (eg on the bias and variance of estimates), without allowing the user to unpick the original disclosive data.	No persons or data items are removed from the census data, and therefore, outputs at national level and high geographies are unaffected by record swapping. At all geographical levels, table totals reflect the sum of the cells and there are no differences between cell counts in different tables. Swapped households have been matched on basic characteristics to preserve data quality and most swapping has been done within the same local authority or middle layer super output area (MSOA). The level of non-response and imputation has a far greater effect on any counts than record swapping. Care was taken to achieve a balance between disclosure risk and data utility. Because records were targeted where the risk of disclosure was greatest, especially people with unusual characteristics who could be recognised in small areas, any analyses based on larger numbers will not be greatly affected. *Source: Statistical Disclosure Control for 2011 Census, Office for National Statistics, 2012*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.38 Describe the impact of any changes in statistical disclosure methods over time. *(Coherence and Comparability)* ■	Users should be informed if there are any changes in statistical disclosure methods that result in incomparable data over time (including historic comparisons).	MICRODATA: Pre-2002, there were no Statistical Disclosure Control (SDC) guidelines on how small cells might represent a disclosure risk in births and deaths statistics. This changed in 2002 to suppress counts smaller than 5. In 2008, the guidelines were amended to indicate cells lower than 3 were suppressed unless relating to stillbirths. For more information see *Disclosure Control Policy for Birth and Death: Briefing Note, Office for National Statistics, 2010.*
B5.39 How well does the output meet user needs after applying statistical disclosure control? *(Relevance)* ■	This should indicate which user needs have been met, and highlight any areas where user needs have not been met (eg on levels of classifications available or geographies) because of statistical disclosure methodologies.	A known gap in our ability to satisfy users' expectations is that we cannot provide sector detail below the level of manufacturing, services and UK Continental Shelf. Unfortunately, due to the relatively small sample size of Quarterly Operating Profits Survey (QOPS), ONS has concerns about both the quality of estimates that could be achieved and the potential disclosure of individual respondents if greater detail were published. *Source: Profitability of UK Companies, Quality and Methodology Information, Office for National Statistics, 2012*
B5.40 Does statistical disclosure control affect significant relationships in the data? *(Accuracy and Reliability)* ■	Any affect of statistical disclosure control on significant relationships in the data should be broadly stated to avoid the risk of users unpicking the data and restoring disclosive data.	At the individual level, record swapping has no impact on the relationships between variables since only geography is swapped between households. The results ~~for the~~ of a Cramer's V test show that over-imputation does have an impact on the relationships between variables. This could potentially impact on the relationship between individuals in households where some variables for some individuals will be imputed and others will not. *Source: Evaluating a statistical disclosure control strategy for 2011 census outputs, Office for National Statistics, 2011*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.41 Have harmonised statistical disclosure control methods been applied? *(Coherence and Comparability)* ■	This enables users to assess whether there are differences between datasets in their disclosure control treatment.	Prior to releasing tabular outputs from the 2001 Census, concerns were raised that the public would perceive that no disclosure control method had been applied. ONS decided that the additional method of small cell adjustment was required for tabular outputs. The small cell adjustments added more uncertainty and removed small cells from tabular outputs. Northern Ireland Statistics and Research Agency (NISRA) also applied the additional method of small cell adjustment but General Register Office for Scotland (GROS) did not. This late change in Statistical Disclosure Control (SDC) methodology and lack of UK harmonisation caused a number of problems for users. In November 2006 the UK SDC Policy position (ONS (2006) for the 2011 Census was agreed by the Registrars General of Scotland, England and Wales and Northern Ireland. The Registrars General agreed to aim for a common UK SDC methodology for 2011 Census outputs to achieve harmonisation. The SDC Policy position is based on the principle of protecting confidentiality set out in the National Statistics Code of Practice. *Source: Evaluating a statistical disclosure control strategy for 2011 census outputs, Office for National Statistics, 2011*
B5.42 When were statistical disclosure control methods last reviewed? Describe the results. *(Relevance)* ■	This gives an indication of the effectiveness of disclosure control methodologies, in showing how new developments (such as new technologies, the availability of additional information or new publication formats) have been accounted for in updated methodologies. The description should include reference to any changes implemented as a result of the review.	In November 2006, the UK Statistical Disclosure Control (SDC) Policy position for the 2011 Census was agreed by the Registrars General of Scotland, England and Wales and Northern Ireland. In July 2007, a review of a wide range of SDC methods was undertaken assessing them against a set of qualitative criteria in line with the policy statement made by the Registrars General. This resulted in three SDC methods being short-listed for further evaluation to assess risk and utility quantitatively, and this short-list was agreed by the UK Census Committee (UKCC) including the Registrars General. This paper provides an evaluation of the three methods. The approach to this evaluation has been reviewed by the UKCDMAC SDC subgroup and agreed with the UK SDC subgroup. The evidence provided in this paper has been used to inform a recommendation for the SDC method to be used for 2011 Census tables. This has resulted in agreement that record swapping will be the primary strategy for disclosure control. *Source: Evaluating a statistical disclosure control strategy for 2011 census tabular outputs, Office for National Statistics, 2011*
B5.43 Have revised outputs been assessed for potential disclosure risks? *(Accessibility and Clarity)* ■	This assures users that any threats to confidentiality in revised datasets or outputs have been assessed and addressed.	The revised datasets have been subject to the same assessment as the original dataset and all identified potential disclosure risks have been addressed with the application of statistical disclosure control methods.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.44 Describe any restrictions on access to the dataset. *(Accessibility and Clarity)* ■	This should inform users of any restrictions on access, and give broad details on the secure environments in place to protect data (including whether secure environments meet British Standard 7799 and, if not, a broad description of the secure environment that is in place).	The ONS Microdata Release Panel (MRP) authorises the re-use of statistical and research data held by ONS. Access to unpublished data can only be provided for research and statistical purposes and must be approved by the MRP. Where there is no other statutory gateway, the customer must apply for Approved Researcher Status, in addition to a request for data being authorised. ONS must also approve release of disclosive data on births and deaths which are provided for research and statistical purposes to the NHS and other health bodies under the section 42 of the Statistics and Registration Service Act. *Source: Disclosure Control Policy for Birth and Death Briefing Note, Office for National Statistics, 2010*
B5.45 Describe any restrictions on the use of the dataset. *(Accessibility and Clarity)* ■	This informs users of any restrictions governing their use of a dataset, (for example, restrictions on releasing the data to third parties, disclaimers that must accompany outputs from the data, contractual agreements limiting use, etc).	The Business Register and Employment Survey (BRES) is conducted under the Statistics of Trade Act (STA) 1947. This Act imposes restrictions on the way that data collected during the survey may be used. The main aim of these restrictions is to protect the identity of individual enterprises that have completed questionnaires, from being disclosed or otherwise deduced. The BRES outputs published on the National Statistics website have already been subjected to the suppression tests and so the issue of confidentiality does not arise. However, the following BRES output has not been suppressed and contains potentially disclosive cells: (i) Employment information extracted by users of the Nomis database. Access to Nomis is restricted, by the provisions of the ETA 4(3)(f), to holders of a Notice. As this output has not been subject to suppression, users of the output (i) above are personally responsible for ensuring that any information which they publish or pass on to other users not named on their Notice, does not contain disclosive data. *Source: BRES 2011: Guide to use of potentially confidential data, Office for National Statistics, 2011*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
Note B5: Notation for Information Loss Measures: B5.45, B5.46, B5.47, B5.48, B5.49, B5.50 *(Accuracy and Reliability)* ■	Information Loss measures are used to measure the statistical changes between an original table and a protected table allowing suppliers of data to compare disclosure control methods. Information loss measures can also be considered from a user perspective to assess quality of tabular outputs after disclosure control has been applied. The general notation for information loss measures is described in the example.	Let *T* be a collection of tables, and *m* the number of tables in the collection. Denote the number of cells in table $t \in T$ as l_t and let n_{tk} be the cell frequency for cell $k \in t$ When used, the superscript *orig* indicates the original table and the superscript *pert* indicates the protected table. When no superscript is applied the formula applies to both tables, original and protected, separately.
B5.46 Distance Metrics. *(Accuracy and Reliability)* ■	Distance metrics are used to measure the statistical distance between an original table and a protected table after disclosure control methods have been applied. The metrics are calculated for each table with the final measure being the average across the tables. A variety of metrics can be used: (1) Hellingers' Distance Hellingers' distance emphasizes changes in small cells and is particularly appropriate when assessing sparse tables. (2) Absolute Average Distance (AAD) AAD is based on the absolute difference between the original and protected cells. This measure is easily interpretable as the average change per cell. (3) Relative Absolute Distance (RAD) RAD is based on the relative absolute difference between the original and protected cell relative to the original cell value. This measure is particularly suitable for assessing tables with a large variability in cell values.	(see information loss measures [see note B5 above] example for general notation) (1) $HD_t = \sqrt{\sum_{k \in t} \frac{1}{2} (\sqrt{n_{tk}^{pert}} - \sqrt{n_{tk}^{orig}})^2}$ next calculate: $\overline{HD} = \frac{1}{m} \sum_{t \in T} HD_t = \frac{1}{m} \sum_{t \in T} \sqrt{\sum_{k \in t} \frac{1}{2} (\sqrt{n_{tk}^{pert}} - \sqrt{n_{tk}^{orig}})^2}$ (2) $AAD = \frac{1}{m} \sum_{t \in T} \frac{\sum_{k \in t} \left\| n_{tk}^{pert} - n_{tk}^{orig} \right\|}{l_t}$ (3) $RAD = \frac{1}{m} \sum_{t \in T} \sum_{k \in t} \frac{\| n_{tk}^{pert} - n_{tk}^{orig} \|}{n_{tk}^{orig}}$

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.47 Standard Error for Distance Metric. *(Accuracy and Reliability)* ■	When computing the average distance metric, the standard error is often used to take into account the dispersion of values of the metric between tables. An example is shown for the Hellingers' distance.	(see information loss measures [see note B5 above] example for general notation) SE (for Hellingers' Distance) = $\sqrt{\frac{1}{m-1}\sum_{t\in T}(HD_t - \overline{HD})^2}$
B5.48 Impact on Totals and Subtotals. *(Accuracy and Reliability)* ■	A user may be interested in changes to the grand total of a variable or changes to the subtotals by the variable categories (eg changes to totals of males and totals of females for the gender variable). The impact is expressed either as a relative absolute distance for a total or average absolute distance per cell.	(see information loss measures [see note B5 above] example for general notation) Let *C(t)* be a defined collection of cells in table *t*, let $l_{C(t)}$ denote the number of cells in the collection *C(t)* and let $n_{C(t)} = \sum_{k\in C(t)} n_{tk}$ Average Absolute distance per Cell: $AAT = \frac{1}{m}\sum_{t\in T}\frac{\left\lvert n_{C(t)}^{pert} - n_{C(t)}^{orig}\right\rvert}{l_{C(t)}}$ Relative Absolute distance for a Total $RAT = \frac{1}{m}\sum_{t\in T}\frac{\left\lvert n_{C(t)}^{pert} - n_{C(t)}^{orig}\right\rvert}{n_{C(t)}^{orig}}$

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.49 Impact on Measures of Association. *(Accuracy and Reliability)* ■	A statistical analysis that is frequently carried out on contingency tables is a test for independence between categorical variables that span the table. The test for independence for a two-way table is based on a Pearson Chi-Squared Statistic. In simple terms, this measures the strength of the relationship between the variables in the table. The measure indicates the percent relative difference on the measure of association (Cramer's V) between the original and perturbed tables.	(see information loss measures [see note B5 above] example for general notation) The Pearson Chi squared statistic for table *t* is: $\chi_t^2 = \sum_{k \in t} \frac{(n_{tk} - e_{tk})^2}{e_{tk}}$ Where $e_{tk} = \frac{n_{t(i\cdot)} \times n_{t(\cdot j)}}{n_t}$ is the expected count for cell *k* in row *i,* denoted $t(i\cdot)$ and column *j,* denoted $t(\cdot j)$ of table *t*. And $n_{t(i\cdot)} = \sum_{k \in t(i\cdot)} n_{tk}$, $n_{t(\cdot j)} = \sum_{k \in t(\cdot j)} n_{tk}$, $n_t = \sum_{k \in t} n_{tk}$ Cramer's V: $CV_t = \sqrt{\frac{\chi_t^2 / n}{\min(r_t - 1, c_t - 1)}}$ Where r_t = number of rows c_t = number of columns, for table *t* $RCV_t = 100 \times \frac{CV_t^{pert} - CV_t^{orig}}{CV_t^{orig}}$

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.50 Impact on Dispersion of Cell Counts. *(Accuracy and Reliability)* ■	This information loss measure compares the dispersion of the cell counts across the tables before and after the disclosure control method has been applied. Note that if the table has been protected via suppression, then the suppressed cells would have to be given imputed values in order to calculate this measure. For example suppressed cells can be given imputed values based on a row average of unsuppressed cells.	(see information loss measures [see note B5 above] example for general notation) For each table *t*, we calculate: $D_t = \frac{1}{l_t - 1}\sum_{k \in t}(n_{tk} - \bar{n}_t)^2$ where $\bar{n}_t = \frac{\sum_{k \in t} n_{tk}}{l_t}$ Next calculate $DR_t = \frac{D_t^{pert}}{D_t^{orig}}$ To obtain a quality measure across several tables we calculate the average of these ratios: $D = \frac{1}{m}\sum_{t \in T} DR_t$
B5.51 Impact on Rank Correlation. *(Accuracy and Reliability)* ■	One statistical tool for inference is the Spearman's Rank Correlation, which tests the direction and strength of the relationship between two variables. The statistic is based on comparing the ranking (in order of size) of the variable values. Therefore, one important assessment of the impact of disclosure control is to compare rankings of the data before and after perturbation. This information loss measure compares the ordering of tables that all have the same structure, for a given cell in that structure. Normally the tables represent different geographical areas. To calculate the measure, group the tables (for example, into deciles) according to the size of the original counts in a selected cell (for example, unemployment). Repeat this process for the perturbed tables, ordering by size in the selected cell and by the original ranking, to maintain consistency for tied values. The information loss measure is the average number of tables that have moved into a different rank group after disclosure control has been applied.	(see information loss measures [see note B5 above] example for general notation) For selected cell k, if table *t* is in a different group after protection has been applied let $I_{tk} = 1$ otherwise let $I_{tk} = 0$ $RC_k = \frac{100 \times \sum_{t \in T} I_{tk}}{m}$

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B5.52 Graphical Displays of Utility. *(Accuracy and Reliability)* ■	A useful approach can be to visualise counts of a particular characteristic by geography, examining the data before and after a methodology has been applied. "Error bands" can be used to highlight geographies (eg wards, output areas, etc which are failing utility criteria). Note: given the use of the disclosive data, this should only be used as an internal technique and not published.	Record swapping was applied to the 2011 Census. The proportion of people in particular ethnic groups by Output Area (OA) was graphed before (x) and after (y) swapping. OAs where the proportion remained unchanged lay on the line x = y. Error bands were used to indicate those OAs that had significant change and that needed further investigation.
B5.53 Is there evidence to suggest the outputs contain 'sufficient uncertainty'? *(Accuracy and Reliability)* ■	This could involve a numerical assessment of the level of risk remaining in the outputs. It is generally impossible to eliminate the disclosure risk entirely without rendering the data useless. Risk should be reduced to an acceptable level.	Record swapping was applied to the 2011 Census. Doubt was used to measure whether sufficient uncertainty had been achieved. Doubt is a theoretical measure devised as a combination of (1) the percentage of real cases of attribute disclosure in a table that are protected, and (2) the percentage of apparent cases of attribute disclosure that are not real. For more information see *Statistical Disclosure Control for 2011 UK Census - Q&A, Office for National Statistics, 2011.*

Beyond the five dimensions of statistical output quality outlined in Table A1, the European Statistical System Handbook for Quality Reports also identifies other considerations which are relevant to measuring and reporting on quality. Further information is included on page 9.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

B6. Dissemination

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B6.1 Describe key user needs for timeliness of data and how these needs have been addressed. *(Timeliness and Punctuality)* ■	A statement should be provided to indicate how timely the data are for specified needs, and how timeliness has been secured, (eg by reducing the time lag to a number of days rather than months for monthly releases).	The quarterly release was expressly designed to fill an information gap and meet the needs of a key user: Race Online 2012. The organisation Race Online 2012 was established to help more people get online for the first time by 2012. Race Online 2012's need for more frequent statistics on people who had never used the Internet, could not be met by the annual Internet Access: Households and Individuals survey. Therefore, ONS started publishing an experimental quarterly update on Internet users and non-users. *Source: Internet Access - Quarterly Update, Quality and Methodology Information, Office for National Statistics, 2012*
B6.2 Time lag from the reference date/period to the release of the provisional output. *(Timeliness and Punctuality)* ■	This indicates whether provisional outputs, where appropriate, are timely with respect to users' needs.	Provisional business investment estimates are published around eight weeks after the end of the reference quarter, on the same day as the National Accounts' Second Estimate of GDP release. *Source: Business Investment, Quality and Methodology Information, Office for National Statistics, 2013*
B6.3 Time lag from the reference date/period to the release of the final output. *(Timeliness and Punctuality)* ■	This indicates whether final outputs are timely with respect to users' needs.	Final figures on marriages are published in the winter/spring each year (just over two years after the end of the reference period). *Source: Marriages in England and Wales, Quality and Methodology Information, Office for National Statistics, 2012*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B6.4 Time lag from the end of data collection to the publication of provisional estimates. *(Timeliness and Punctuality)* ▲	This indicates whether provisional estimates are timely with respect to users' needs.	The Internet Access data are collected by the Opinions Survey in January, February and March. This is over a 19 week period, including four weeks for developing and testing the questionnaire in BLAISE (a computer assisted survey design and processing system), producing showcards when required and preparing derived variables. During the final five weeks of the survey period, the data are cleaned, weighted, analysed and tabulated by the research team. The estimates are published in the Internet Access Statistical Bulletin on the ONS website in August of the year that the data are collected. EU aggregates are published around the end of the year on the Eurostat website. *Source: Internet Access - Households and Individuals, Quality and Methodology Information, Office for National Statistics, 2012*
B6.5 Delay to accessing / receiving data from the administrative source. *(Relevance)* ●	This indicator provides information on the proportion of the time from the end of the reference period to the publication date that is taken up waiting to receive the administrative data. This is calculated as a proportion of the overall time between reference period and publication date to provide comparability across statistical outputs. Only the final dataset used for the statistical output should be considered. If a continuous feed of data is received, the 'last' dataset used to calculate the statistical output should be used in this indicator. If multiple administrative sources are being used, the indicator can be calculated for each source and then averaged, potentially weighting by the source's contribution to the final estimate.	$\frac{\text{Time from the end of reference period to receiving Admin data}}{\text{Time from the end of reference period to publication date}} \times 100\%$ Information should be provided on what the statistical producer is required to do with the data between receiving it and publication, to inform users of whether the delay is detrimental to the quality of the output.
B6.6 Time lag between scheduled release date and actual release date. *(Timeliness and Punctuality)* ■	This is a measure of the punctuality of the output. If appropriate, reasons behind the delay in publication should be provided, along with information on how users were informed of the delay.	Publication of Blue Book 2011 was delayed. Key users were informed by direct email as soon as it was clear there would be a delay to the publication date. All users were informed via this webpage: http://www.ons.gov.uk/ons/media-centre/statements/national-accounts-timetable-sep-2011/national-accounts-publication-timetable.html which was published over a month ahead of the scheduled Blue Book publication date. The delay for most Blue Book chapters was from 1st November 2011 to 23rd November 2011 but Chapter 2 was still published on time on 1st November.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B6.7 For new and ad hoc surveys, time lag from the commitment to undertake the survey to the release date. *(Timeliness and Punctuality)* ▲	This indicates whether outputs from new or ad hoc surveys are timely with respect to users' needs. The length of time required to design and implement a survey through to the provision of statistical outputs depends upon the scale and complexity of a given information requirement.	The ONS Opinions and Lifestyle Survey offers a 14 week turn around from the specification of an information requirement, to the delivery of high quality statistical outputs. However, new surveys can take longer to turn-around, such as, the European Health Interview Survey (EHIS). EHIS is a new survey undertaken by ONS on behalf of Eurostat. It is sponsored by four government bodies in the UK, with commitment for undertaking the survey given in October 2012. Eurostat require the delivery of data in September 2015, providing a three year lag time between commitment to undertake the survey, to delivery of final outputs.
B6.8 Frequency of production. *(Timeliness and Punctuality)* ■	This indicates how timely the outputs are, as the frequency of publication indicates whether the outputs are up to date with respect to users' needs.	In order to provide timely data to users, key headline figures for birth statistics are provided quarterly as provisional data. These data are then finalised once the annual data are published. *Source: Birth Statistics, Quality and Methodology Information, Office for National Statistics, 2012*
B6.9 Timetable for release of data. *(Accessibility and Clarity)* ■	Providing timetables of future releases enables equal access to outputs for all groups of users. A link should be provided to the release timetable, available on the Publication Hub. As well as regular releases, there should be a process for making data from ad-hoc requests widely available.	For more details on related releases, the UK National Statistics Publication Hub is available online and provides 12 months' advanced notice of release dates: http://www.statistics.gov.uk/hub/index.html If there are any changes to the pre-announced release schedule, public attention will be drawn to the change and the reasons for the change will be explained fully at the same time, as set out in the Code of Practice for Official Statistics. *Source: Long-term International Migration estimates, Quality and Methodology Information, Office for National Statistics, 2013*

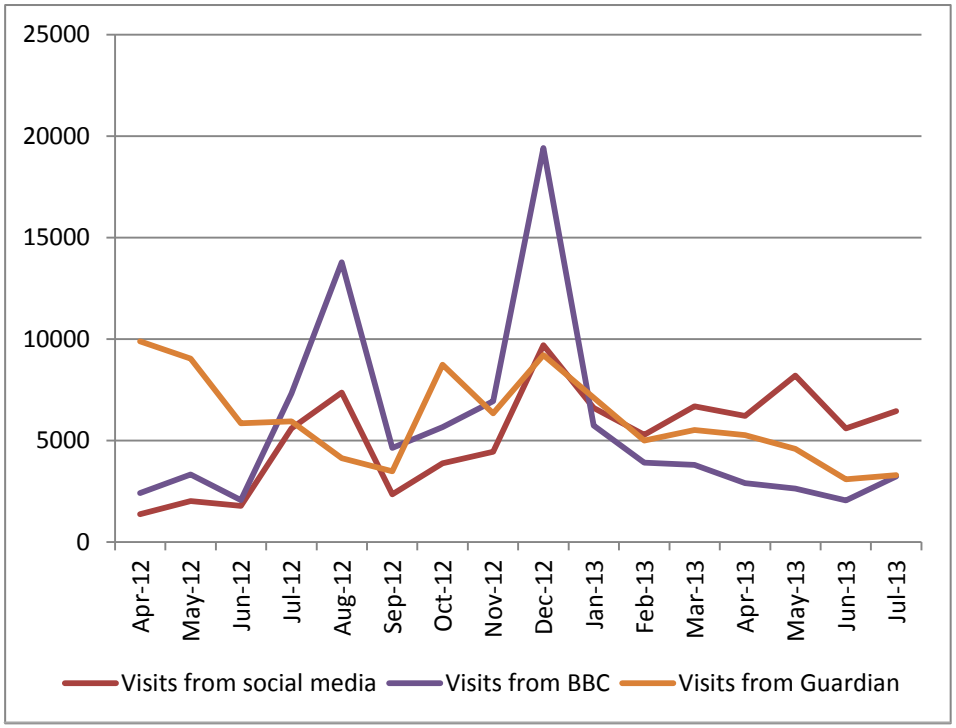
▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B6.10 Describe the range of statistical products that are available to meet user needs. *(Accessibility and Clarity)* ■	Information should be provided on all statistical products that are produced for the output, and, if possible, how the different types of products meet the needs of the different user groups. Examples of statistical products could include: a statistical bulletin; a YouTube clip; a Facebook entry; a Twitterfeed; a number of short stories / analytical pieces on the website; a quality report published on our website; detailed underlying tables on our website; data put on NOMIS; identifiable data made available through VML / secure data service; and occasional technical articles.	Labour market estimates are published every month and include tables, text and charts. The data are widely available, generally free of charge, through a range of media. Labour Market Statistical Bulletins, associated data tables and Time Series Datasets are available to download, free of charge, from the National Statistics website. The Labour Market Statistical Bulletins contain additional textual analysis and charts that supplement the data in the tables. A highly disaggregated dataset, which covers a wealth of data for local areas, is available free from the NOMIS website. http://www.nomisweb.co.uk/ *Source: Summary Quality Report for Labour Market Data Releases, Office for National Statistics, 2011*
B6.11 Document availability of quality information. *(Accessibility and Clarity)* ■	Users should be informed of what quality information is available to help them judge the strengths and limitations of the output in relation to their needs.	A *Quality and Methodology Information Report for ILCH* (148.9 Kb Pdf) is available. This report describes in detail, the intended uses of the statistics presented in this publication, their general quality and the methods used to produce them. *Source: Index of Labour Costs per Hour (ILCH) Q4 2012 (experimental) - Background notes*
B6.12 Are there links to metadata? *(Accessibility and Clarity)* ■	Metadata help users to make appropriate use of the data. Links to available metadata ensure that this information is accessible. Users should also be made aware of and provided with links to other articles and related outputs, including more detailed methodology papers.	For more information on data quality, legislation and procedures relating to conception statistics see *Conceptions Metadata 2011 (123.6 Kb Pdf)* *Source: Conceptions in England and Wales, 2011 - Further information, Office for National Statistics, 2013*
B6.13 Describe procedures to request access to micro datasets, including costs. *(Accessibility and Clarity)* ■	Procedures to request access to micro datasets should be clearly described and any cost implications detailed.	ONS provides two routes to getting access to data: one for controlled data, through the Virtual Microdata Laboratory (VML), the other for safeguarded data, which can be sent securely to users. The processes required and their cost implications are set out on the ONS website. The VML section sets out the steps of the process, and allows users to download the necessary forms and documentation. The 'access to ONS data' pages likewise set out the process for users, providing contact details and the necessary forms for the approved researcher process.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B6.14 Describe procedures to request access to non-published data. *(Accessibility and Clarity)* ■	Procedures to request access to unpublished data should be clearly described, and the average time and cost of providing simple tables of non-published data should be detailed.	**Data Advice and Relations Team (DART) -** are a dedicated service team who offer a chargeable service for bespoke tables using Social Survey data. • An external customer can contact the team to set up bespoke analysis for them using Labour Force Survey, Annual Population Survey, International Passenger Survey and the General Lifestyle Survey datasets. • These requests usually drill down to the lower levels of analysis from which is supplied in the main publications (eg Labour market produce UK/regional estimates, we can supply data at a Local Authority level). • Once the customer's specification is clear, a quote is provided for the work using the ONS Charging policy. • On average, requests usually take a day to complete but more complex data requests can take longer. • Once work is completed for the customer, it is then decided whether the data should be published or listed on the ONS website. • When deciding whether to list or publish analysis, the Ad Hoc Request Guidance document on the intranet is used for guidance. • If it is decided to publish the tables, everyone can access the data as it is published on the ONS website. • If it is decided to list the tables, the requestor for this data would need to contact the ONS customer contact centre for them to supply.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B6.15 Where relevant, provide details of web and social media metrics. *(Accessibility and Clarity)* ■	This is an indicator of the accessibility of the statistics on the website. In addition, it may denote the relevance of the output and associated products to users' needs. One potential measure is the number of clicks required to access the statistics from the homepage.	The graph below denotes the number of visits to the ONS website by users coming from social media. 
B6.16 Detail contact points for further information, including technical information. *(Accessibility and Clarity)* ■	Users should be provided with details of contacts who are able to provide them with further information.	General enquiries on the Services Producer Price Indices (SPPI) series, compilation methods, developmental articles, quality information or difficulties in finding the latest figures can be emailed to the SPPI team in ONS at sppi@ons.gov.uk. *Source: Services Producer Price Indices (SPPI), Quality and Methodology Information, office for National Statistics, 2012*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B6.17 Describe any differences between domains. *(Coherence and Comparability)* ■	Differences between domains may be found in any relevant characteristic of the domain, or in the methods used, including: definitions; target and study populations; sampling unit; sampling methods; data collection methods; processing methods; and data analysis methods. These differences will impact on data comparability between domains.	Four different sampling frames are used in the UK Labour Force Survey. Great Britain is split into two areas: south of the Caledonian Canal, comprising all of England, Wales and most of Scotland; and north of the Caledonian Canal in Scotland. Northern Ireland has its own sampling frame. A separate list of NHS accommodation in Great Britain is maintained. For most of Great Britain, the sampling frame is the Royal Mail's Postcode Address File (PAF), a database of all addresses receiving mail. The list is limited to addresses receiving relatively few items of post per day, so as to exclude businesses. Because of the very low population density in the far north of Scotland (north of the Caledonian Canal), telephone directories are used as sampling frames, and interviews are carried out by telephone because face-to-face interviews would be too expensive. In Northern Ireland POINTER, which is the government's central register of domestic properties, is used. *Source: Labour Force Survey (LFS), Quality and Methodology Information, Office for National Statistics, 2011*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B6.18 Estimate mean absolute revision between provisional and final statistics. *(Accuracy and Reliability)* ■	This provides an indication of reliability (which is one aspect of accuracy) in that provisional statistics may be based on less information than is available for final statistics, and may have been processed more quickly. However, the value of provisional statistics lies in their timely provision. Statistics that are not revised are not necessarily more accurate than those that are revised often: it may be simply due to different revisions policies. The number of estimates from which the mean absolute revision is calculated, and the gap between the 'preliminary' and 'later' estimates should be predefined and should be unchanged between publications, in order to be open and clear. This indicator should be based upon as many previous observations as possible, in order for the results to be sufficiently reliable (for example, it is recommended that the indicator should be based on at least 20 observations for quarterly outputs). The mean absolute revision should be expressed in the same units as the estimate (eg percentage points, £m etc). Commentary should be provided in terms of what the revisions mean in terms of the quality of the data, including a comparison with revisions made in previous periods if appropriate. Any known reasons for revisions between provisional and final estimates should be described, along with reasons for particularly high or low revisions over time. Note that there is a difference between mean revisions across series (eg across industries) and mean revisions across time. The most appropriate indicator will be dependent on the output itself.	In the first quarter of 2012, net investment has been revised from £34.0 billion to £37.6 billion, partly as a result of late questionnaires being received and partly as a result of disaggregate data revisions. Revisions to data provide one indication of the reliability of key indicators. The table below compares the first published estimate for total net investment with the equivalent figure published three years later. The data starts with the first estimate published for Q3 2004 (in December 2004) and compares this with the estimate for the same quarter published three years later (in December 2007). The difference between these two estimates is calculated and this process is repeated for five years of data (all quarters up to Q2 2009). The averages of this difference (with and without regard to sign) are shown in the right hand columns of the table. These can be compared with the value of the estimate in the latest quarter. A statistical test has been applied to the average revision to find out if it is statistically significantly different from zero. An asterisk (*) shows if the test is significant. **Table 4: Revisions between first publication and estimates three years later** £ billion **Value in latest quarter** / **Average revision** / **Average revision without regard to sign** **Total net investment** 20.6 1.3 5.6 *Source: MQ5: Investment by Insurance Companies, Pension Funds and Trusts, Q2 2012, Office for National Statistics, 2012*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B6.19 Estimate the mean revision between provisional and final estimates. *(Accuracy and Reliability)* ■	This is an indicator of bias. It shows whether revisions are being made consistently in one direction - ie if early estimates are consistently under or over estimating the later, more information-based figures. The number of estimates from which the mean revision is calculated and the gap between the 'preliminary' and 'later' estimates should be predefined and should be unchanged between publications, in order to be open and clear. This indicator should be based upon as many previous observations as possible, in order for the results to be sufficiently reliable (for example, it is recommended that the indicator should be based on at least 20 observations for quarterly outputs). A t-test can be used to assess whether the mean revision is statistically different from zero. An adjusted t-test may need to be used if revisions for different reference periods are correlated. The mean absolute revision should be expressed in the same units as the estimate (eg percentage points, £m etc). Commentary should be provided in terms of what the revisions mean in terms of the quality of the data, including a comparison with revisions made in previous periods if appropriate. Any known reasons for revisions between provisional and final estimates should be described, along with reasons for particularly high or low revisions over time. Note: There is a difference between mean revisions across series (eg across industries) and mean revisions across time. The most appropriate indicator will be dependent on the output itself.	See above (B6.18) for example.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B6.20 Flag data that are subject to revision and data that have already been revised. *(Accuracy and Reliability)* ■	This indicator alerts users to data that may be, or have already been, revised. This will enable users to assess whether provisional data will be fit for their purposes.	Estimates for the most recent quarters are provisional and are subject to revision in the light of updated source information. ONS currently provides an analysis of past revisions in the Gross Domestic Product (GDP) and other Statistical Bulletins which present time series. ONS has a webpage dedicated to revisions to economic statistics which brings together ONS work on revisions analysis, linking to articles, revisions policies and key documentation from the Statistics Commission's report on revisions. Revisions to data provide one indication of the reliability of key indicators. *Source: Quarterly National Accounts, Statistical Bulletin, Office for National Statistics, Q1 2013*
B6.21 Provide timetable for revisions. *(Accessibility and Clarity)* ■	Information on when revisions will be available enables users to judge whether the release of revisions will be timely enough for their needs and whether revisions are delivered punctually.	The 2011 edition of the UK National Accounts - the "Blue Book" - will contain significant changes: new industrial and product classifications, required by European Regulation (SIC07 and CPA08, respectively), improved methods of deflation and some additional improvements in the financial services area. (These were set out in UK National Accounts, Content of Blue Book, 25 March 2011 (46.8 Kb Pdf). *Source: Methods Changes in the 2011 Blue Book, Office for National Statistics, 2011*
B6.22 For ad hoc revisions, detail revisions made and provide reasons. *(Accuracy and Reliability)* ■	Where revisions occur on an ad hoc basis, this may be because earlier estimates have been found to be inaccurate. Users should be clearly informed of the revisions made and why they occurred. Clarifying the reasons for revisions guards against any misinterpretation of why revisions have occurred, while at the same time making processes (including any errors that may have occurred) more transparent to users.	As announced on 20 June 2013, publication of the detailed industry breakdowns normally available in the Business Investment Statistical Bulletin will be delayed until 31 July 2013. However, Business Investment release will trial a new approach, and will include sector and asset breakdowns of Gross Fixed Capital Formation (GFCF) not previously available in the release. Industry breakdowns will be re-instated as part of the regular release from August 2013. *Source: Impact of changes in the National Accounts and economic commentary for Q1 2013*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

<table>
<tr><th>Quality Measure / Indicator and Quality Theme</th><th>Description</th><th>Example / Formula</th></tr>
<tr>
<td>B6.23 Reference/link to detailed revisions analyses.

(Accessibility and Clarity)</td>
<td>Users should be directed to where detailed revisions analyses are available.</td>
<td>Table 6 shows the revisions to month 1 and month 2 estimates of GDP. The analysis of revisions between month 1 and month 2 uses month 2 estimates published from August 2008 (Q2 2008) to May 2013 (Q1 2013). The analysis of revisions between month 2 and month 3 uses month 3 estimates published from June 2008 (Q1 2008) to March 2013 (Q4 2012).

Table 6: Revisions to early estimates of GDP growth
<table>
<tr><th rowspan="2">Revisions to GDP growth</th><th rowspan="2">GDP Growth in the latest period %</th><th colspan="2">Revisions between early estimates of GDP growth (quarterly, CVM)</th></tr>
<tr><th>Average over the last five years</th><th>Average over the last five years without regard to sign (average absolute revision)</th></tr>
<tr><td>Between M1 and M2</td><td>0.3</td><td>-0.02</td><td>0.06</td></tr>
<tr><td>Between M2 and M3</td><td>0.3</td><td>-0.06</td><td>0.11</td></tr>
</table>
An article titled 'Updated analysis: Why is GDP revised?' (300.9 Kb Pdf) published on 13 June 2013, is available on the ONS website.

Source: Background Notes 11, Statistical Bulletin: Quarterly National Accounts, Q1 2013</td>
</tr>
</table>

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B6.24 Compare estimates with other estimates on the same theme and state whether they have been produced in ways that are coherent? *(Coherence and Comparability)* ■	Users should be advised as to whether outputs on the same theme have been produced in comparable ways, or if there are differences in their production, eg in definitions, classifications, sampling methods, data collection methods, processing or analysis. The level of detail provided may depend on whether the outputs have been produced within the same department or produced in other departments and/or countries. Any known reasons for lack of coherence should be given.	In addition to the Retail Sales Index (RSI), ONS also produces retail sales statistics as part of the Annual Business Survey (ABS). There are differences between the RSI and ABS retail figures for the following reasons: the ABS collects and produces statistics on the United Kingdom whereas RSI covers Great Britain, resulting in differences in total sales values; • reporting periods can vary, the ABS allows respondents to provide information on either a calendar year or financial calendar year. In comparison the RSI asks respondents to provide information based on a 4 - 4 - 5 week reporting period or calendar monthly; • for the food sector, the ABS includes automotive fuel sales by supermarkets, for the RSI these are extracted and used for the calculation of the automotive fuel series; and • for chemists, the ABS includes National Health Service receipts, these are excluded from the RSI. *Source: Retail Sales Index, Quality and Methodology Information, Office for National Statistics, 2012*

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Quality Measure / Indicator and Quality Theme	Description	Example / Formula
B6.25 Identify known gaps between key user needs, in terms of coverage and detail and current data. *(Relevance)* ■	Data are complete when they meet user needs in terms of coverage and detail. This indicator allows users to assess, when there are gaps, how relevant the outputs are to their needs.	A known gap in our ability to satisfy users' expectations is that we cannot provide sector detail below the level of manufacturing, services and UK Continental Shelf. Unfortunately, due to the relatively small sample size of Quarterly Operating Profits Survey (QOPS), ONS has concerns about both the quality of estimates that could be achieved and the potential disclosure of individual respondents if greater detail were published. *Source: Profitability of UK Companies, Quality and Methodology Information, Office for National Statistics, 2012*
B6.26 Present any reasons for gaps between user needs and what is produced. *(Relevance)* ■	This indicator gives the main reasons why an output is not relevant to user needs. For example, the study population may not be the same as the user's target population, or the operational definition of a concept may not capture the user's required definition of a concept.	The output may not satisfy all users' requirements for ward-level data broken down into sub-groups based on ethnicity. This is because, in many wards, cell numbers were small and potentially disclosive when using the Level 2 Ethnic Grouping (NS interim standard classification of ethnic groups). Therefore, the broader Level I Ethnic Grouping classification was used for ward-level data. In addition, where cell counts were small and potentially disclosive, even when using this broader ethnic classification, data values were suppressed using primary and secondary suppression.
B6.27 Describe any other adjustments applied to increase the relevance of outputs. *(Relevance)* ■	The relevance of an output may be increased if adjustments to the data are made, eg to weight sub-groups to their population totals, standardisation or to separate seasonal effects from underlying trends. While improving the usefulness of data, these adjustments must be described for transparency and to aid users' understanding of potential data uses and limitations.	To compare levels of mortality between different areas, it is helpful to remove the effects of varying population structures among these areas. This can be done by using a standard set of sex and age-specific mortality rates and applying these to the population of each area. This will produce the number of deaths expected if the standard rates had applied in the area. This expected number is then compared with the number of deaths actually observed in the area.

Beyond the five dimensions of statistical output quality outlined in Table A1, the European Statistical System Handbook for Quality Reports also identifies other considerations which are relevant to measuring and reporting on quality. Further information is included on page 9.

▲ = relevant to survey data based outputs; ● = relevant to administrative data based outputs; ■ = relevant to outputs involving either survey data, administrative data or both

Annex A: Detailed formulas relating to indicator B3.9: Unit response rate by sub-groups.

Formula for household surveys.

Overall response rate;
unweighted:

$$\frac{I + P}{(I+P) + (R+NC+O) + e_c UC + e_n UN}$$

Full response rate, unweighted:

$$\frac{I}{(I+P) + (R+NC+O) + e_c UC + e_n UN}$$

Co-operation rate, unweighted:

$$\frac{I + P}{(I+P) + R + O + e_c UC}$$

Contact rate, unweighted:

$$\frac{(I+P) + R + O + e_c UC}{(I+P) + (R+NC+O) + e_c UC + e_n UN}$$

Refusal rate, unweighted:

$$\frac{R}{(I+P) + (R+NC+O) + e_c UC + e_n UN}$$

Eligibility rate, unweighted:

$$\frac{(I+P) + \left(R+NC+O\right) + e_c UC + e_n UN}{(I+P) + (R+NC+O) + (UC+UN) + NE}$$

Where:
I = complete interview; P = partial interview; R = refusal; NC = non-contact; O = other non-response; NE = not eligible; e_c = estimated proportion of contacted cases of unknown eligibility that are eligible; UC = unknown eligibility, contacted case; e_n = estimated proportion of non-contacted cases of unknown eligibility that are eligible; UN = unknown eligibility, non-contact.

Weighted rates are derived by applying the design weight, π^{-1}, to each variable when calculating the above rates. For example, 'I' becomes the weighted number of complete interviews, 'NC' becomes the weighted number of non-contacts, etc.

Formula for business surveys.

Overall response rate:

Unweighted:

$$\frac{\text{FC + FP + PC + PP}}{\text{(FC+FP+PC+PP) + RNU +NR + e(U)}}$$

Weighted:

$$\frac{\sum_{i \in R_O} w_i x_i}{\sum_{i \in S} w_i x_i}$$

Full response rate:

Unweighted:

$$\frac{\text{FC}}{\text{(FC+FP+PC+PP) + RNU + NR + e(U)}}$$

Weighted:

$$\frac{\sum_{i \in R_F} w_i x_i}{\sum_{i \in S} w_i x_i}$$

Non-response rate:

Unweighted:

$$\frac{\text{RNU + NR + e(U)}}{\text{(FC+FP+PC+PP) + RNU + NR + e(U)}}$$

Weighted:

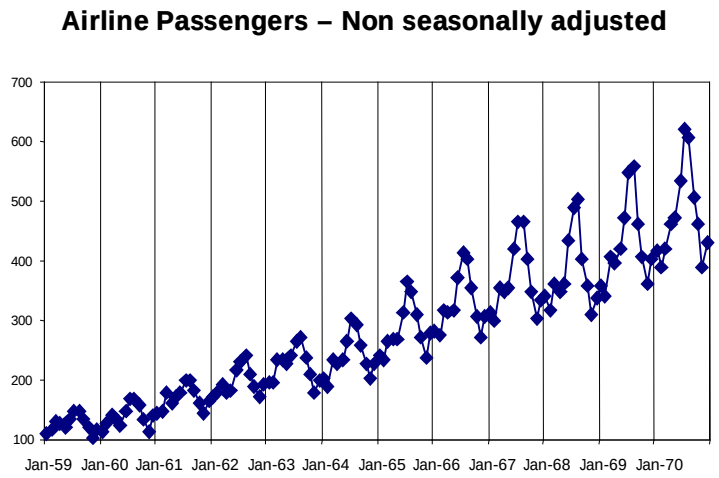
$$\frac{\sum_{i \in Nr} w_i x_i}{\sum_{i \in S} w_i x_i}$$

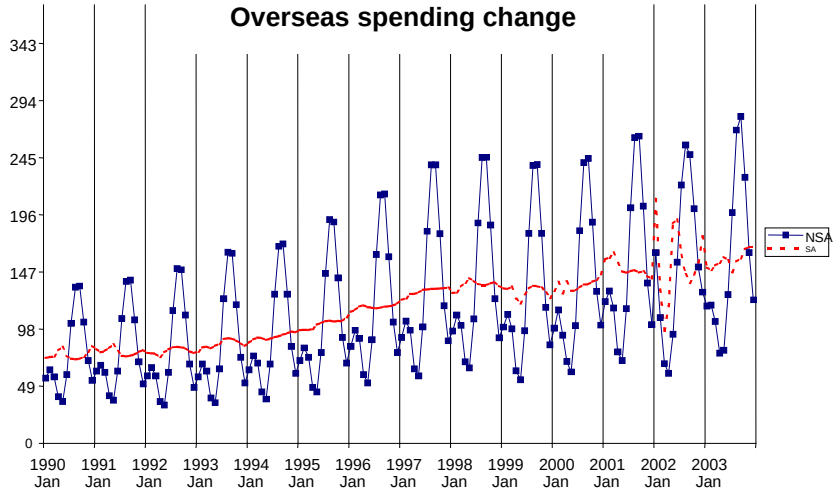
Where: FC = full period return with complete data; FP = full period return with partial data; PC = part period return with complete data; PP = part period return with partial data; RNU = returned but not used; NR = non-response; U = unknown eligibility; I = ineligible (out of scope); e = estimated proportion of cases of unknown eligibility that are eligible. e can be estimated as: $\text{e} = \frac{\text{(FC+FP+PC+PP) + RNU + NR}}{\text{(FC+FP+PC+PP) + RNU + NR + I}}$.

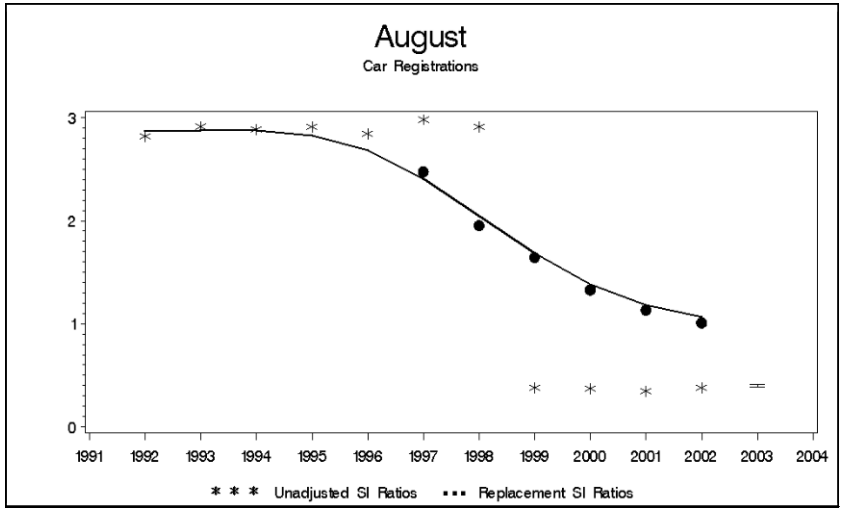
w_i = weight for unit i; x_i = value of auxiliary variable for unit i; R_O = set of all responders; S = set of all eligible sampled units; R_F = set of full responders; Nr = set of non-responders (note that $Nr = S - R_O$).

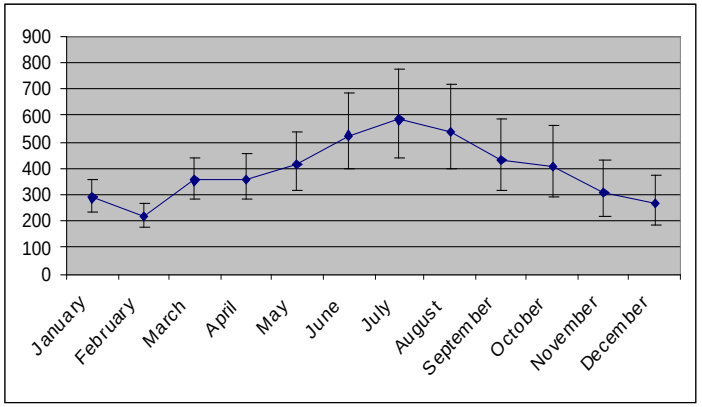
Annex B: Possible techniques to visualise time series analysis

Visualisation of time series can be useful in providing information about the effectiveness of the time series analysis. The following guidance may be useful.

Visualisation	Description	Example
Annex B.1 Non-seasonally adjusted data visual check.	Graphed data can be used in a visual check for the presence of seasonality, decomposition type model (multiplicative or additive), extreme values, trend breaks and seasonal breaks.	The graph below shows the Airline Passengers original series. From the graph it is possible to see that the series has repeated peaks and troughs which occur at the same time each year. This is a sign of seasonality. As the size of the seasonal peaks and troughs are not independent of the level of the trend, a multiplicative decomposition model is appropriate for the seasonal adjustment. This series does not show any particular discontinuities, such as: • outliers – one-off extreme (high or low) values; • level shifts – sudden and sustained changes in the level of the series (shift up or down); or • seasonal breaks – sudden and sustained changes in the size of seasonal peaks (independent of the trend). 

Visualisation	Description	Example
Annex B.2 Compare the non-seasonally adjusted and seasonally adjusted data.	By graphically comparing the non-seasonally adjusted and seasonally adjusted series, the quality of the seasonal adjustment can be assessed. Problems with extreme values, trend breaks or seasonal breaks will be seen as residual seasonality in the seasonally adjusted series.	By comparing the non-seasonally adjusted and the seasonally adjusted series in the graph below, it can be seen that the seasonal adjustment performs well until January 2002 where the presence of an outlier distorts the pattern of the series, both at January 2002 and before and after this point. If the outlier had been correctly treated, the effect would have been isolated to January 2002 and there would have been no residual effects before and after. 

Visualisation	Description	Example
Annex B.3 Graph of the seasonal-irregular series.	It is possible to identify a seasonal break by a visual inspection of the seasonal-irregular series. Any changes in the seasonal pattern indicate the presence of a seasonal break. Other methods include inspection of: period-period growths in the non-seasonally adjusted series, seasonal factors, monitoring and quality statistics.	In the example below, there has been a sudden drop in the level of the seasonal-irregular component for August between 1998 and 1999. This is caused by a seasonal break in the car registration series which was due to the change in the car number plate registration legislation. Permanent prior adjustments should be estimated to correct for this break. (Permanent prior adjustments do not have their effects removed after the seasonal adjustment. They are used for removing effects that should not appear in the seasonally adjusted series). If no action is taken to correct for the break, some of the seasonal variation will remain in the irregular component resulting in residual seasonality in the seasonally adjusted series. The result would be errors in the seasonally adjusted series and increased revisions misleading user interpretation. 

Visualisation	Description	Example
Annex B.4 Graph of the confidence intervals.	Time to be plotted along the x-axis, and forecast estimates and confidence intervals along the y-axis.	

References

Beaumont J F and Mitchell C (2002) *A system for variance estimation due to non-response and imputation (SEVANI)* available at: http://www5.statcan.gc.ca/bsolc/olc-cel/olc-cel?catno=11-522-X20020016729&lang=*eng*

BLUE-ETS project (see specifically Deliverable 4.3) available at: http://www.blue-ets.istat.it/index.php?id=7

ESSnet Admin Data Information Centre available at: http://essnet.admindata.eu/
Eurostat (2008) *NACE Rev.2 Statistical classification of economic activities in the European Community:* http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-RA-07-015

Eurostat: ESS Handbook for Quality Reports 2009 edition:
http://epp.eurostat.ec.europa.eu/portal/page/portal/lang-en/ver-1/quality/documents/EHQR_FINAL.pdf
An updated version is due to be published in late 2013

Generic Statistical Business Process Model:
http://www1.unece.org/stat/platform/display/metis/Generic+Statistical+Business+Process+Model

Information Commissioner's Office (2012) *Anonymisation: managing data protection risk code of practice: http://www.ico.org.uk/news/latest_news/2012/~/media/documents/library/Data_Protection/Practical_application/anonymisation_code.ashx*

Kozlowski L T (1986) *Pack size, reported smoking rates and public health, American Journal of Public Health*, 76 (11) pp 1337-8 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1646752/

Lewis D and Woods J (2013) *Issues to consider when turning to the use of administrative data: the UK experience.* Paper presented at the seventh NTTS (New Techniques and Technologies for Statistics) conference, Brussels, Belgium, 5 to 7 March 2013: http://www.cros-portal.eu/sites/default/files/NTTS2013fullPaper_142.pdf

McLaren C (2009) *Changes to the retail sales methodology Economic and Labour Market Review, 3 (6):* http://www.ons.gov.uk/ons/guide-method/method-quality/specific/economy/retail-sales/changes-to-retail-sales-methodology.pdf

Office for National Statistics (2003) *Report on research into revising internal migration estimates:* http://www.ons.gov.uk/ons/rel/migration1/nhscr-/report-on-research-into-revising-internal-migration-estimates/internal-migration--nhscr----report-on-research-into-revising-internal-migration-estimates.pdf

Office for National Statistics (2004) *Update following further investigation into revising internal migration estimates:* http://www.ons.gov.uk/ons/rel/migration1/nhscr-/update-following-further-investigation-into-revising-internal-migration-estimates/revising-internal-migration-estimates.pdf

Office for National Statistics (2006) *Estimation of Standard Errors for the UK Index of Production, Survey Methodology Bulletin, Issue 59,* available at: http://www.ons.gov.uk/ons/guide-method/method-quality/survey-methodology-bulletin/smb-59/index.html

Office for National Statistics (2007) *UK Standard Industrial Classification 2007 (UK SIC 2007):* http://www.ons.gov.uk/ons/guide-method/classifications/current-standard-classifications/standard-industrial-classification/index.html

Office for National Statistics (2009) *Deaths involving Clostridium difficile:* http://www.ons.gov.uk/ons/rel/subnational-health2/deaths-involving-clostridium-difficile/2009/index.html

Office for National Statistics (2010) *Briefing Note: ONS policy on protecting confidentiality within birth and death statistics (Revised 2010): http://www.ons.gov.uk/ons/guide-method/best-practice/disclosure-control-policy-for-birth-and-death-statistics/disclosure-control-policy-for-birth-and-death--briefing-note.pdf*

Office for National Statistics (2010) *General Lifestyle Survey Technical Appendices 2010: http://www.ons.gov.uk/ons/rel/ghs/general-lifestyle-survey/2010/technical-appendices-2010.pdf*

Office for National Statistics (2010) *Travel Trends, ONS, Appendix E:* http://www.ons.gov.uk/ons/rel/ott/travel-trends/2010/travel-trends---2010.pdf

Office for National Statistics (2011) *2011 Census Security: Report of the Independent Review Team: http://www.ons.gov.uk/ons/guide-method/census/2011/the-2011-census/the-2011-census-project/commitment-to-confidentiality/2011-census--independent-information-assurance-review.pdf*

Office for National Statistics (2011) *Annual Business Survey (ABS) Questionnaire Review - A qualitative study: http://www.ons.gov.uk/ons/guide-method/method-quality/specific/business-and-energy/annual-business-survey/quality-and-methods/abs-questionnaire-review-2011.pdf*

Office for National Statistics (2011) *A Quick Guide to the Retail Sales Index:* http://www.ons.gov.uk/ons/guide-method/method-quality/specific/economy/retail-sales/rsi-quick-guide.pdf

Office for National Statistics (2011) *Conceptions in England and Wales Statistical Bulletin* available at: *http://www.ons.gov.uk/ons/rel/vsob1/conception-statistics--england-and-wales/2011/2011-conceptions-statistical-bulletin.html#tab-Further-information*

Office for National Statistics (2011) *Evaluating a Statistical Disclosure Control (SDC) strategy for 2011 Census outputs: http://www.ons.gov.uk/ons/guide-method/census/2011/the-2011-census/producing-and-delivering-data/confidentiality/evaluating-a-statistical-disclosure-control--sdc--strategy-for-2011-census-outputs.pdf*

Office for National Statistics (2011) *Labour Force Survey User Guide - LFS Background and Methodology, Volume 1: http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/labour-market-statistics/volume-1---2011.pdf*

Office for National Statistics (2011) *Measuring the accuracy of the Retail Sales Index, Economic & Labour Market Review, pp 88-94. http://www.ons.gov.uk/ons/rel/elmr/economic-and-labour-market-review/no--2--february-2011/economic---labour-market-review.pdf*

Office for National Statistics (2011) *Methods Changes in the 2011 Blue Book*: http://www.ons.gov.uk/ons/guide-method/method-quality/specific/economy/national-accounts/articles/methods-changes-in-the-2011-blue-book.pdf

Office for National Statistics (2011) *Statistical Disclosure Control for 2011 Census: http://www.ons.gov.uk/ons/guide-method/census/2011/the-2011-census/processing-the-information/statistical-methodology/statistical-disclosure-control-for-2011-census.pdf*

Office for National Statistics (2011) *Statistical Disclosure Control for 2011 UK Census – Q&A: http://www.ons.gov.uk/ons/guide-method/census/2011/the-2011-census/processing-the-information/statistical-methodology/statistical-disclosure-control-for-2011-uk-census---q-a.pdf*

Office for National Statistics (2012) *Consumer Prices Index and Retail Prices Index: Updating Weights for 2012* available at: *http://www.ons.gov.uk/ons/rel/cpi/cpi-and-rpi-index--updating-weights/updating-weights--2012/cpi-and-rpi-index.pdf*

Office for National Statistics (2012) *Consumer Price Indices – Technical Manual*: http://www.ons.gov.uk/ons/guide-method/user-guidance/prices/cpi-and-rpi/consumer-price-indices-technical-manual.pdf

Office for National Statistics (2012) *General Lifestyle Survey Overview – A report on the 2010 General Lifestyle Survey:* http://www.ons.gov.uk/ons/rel/ghs/general-lifestyle-survey/2010/general-lifestyle-survey-overview-report-2010.pdf

Office for National Statistics (2012) *Index of Labour Costs per Hour (ILCH) Q4 2012 (experimental) Statistical Bulletin* available at: *http://www.ons.gov.uk/ons/rel/ilch/index-of-labour-costs-per-hour--experimental-/q4-2012/stb-index-of-labour-cost-per-hour-q4-2012--experimental-.html#tab-background-notes*

Office for National Statistics (2012) *MQ5: Investment by Insurance Companies, Pension Funds and Trusts, Q2 2012 Statistical Bulletin* available at: *http://www.ons.gov.uk/ons/rel/fi/mq5--investment-by-insurance-companies--pension-funds-and-trusts/q2-2012/pdf-mq5-q2-2012.pdf*

Office for National Statistics (2012) *User Information Guide BRES 2012: Guide to use of potentially confidential data: https://www.nomisweb.co.uk/notices/docs/infoguide.pdf*

Office for National Statistics (2013) *Beyond 2011 Matching Anonymous Data* http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/reports-and-publications/beyond-2011-matching-anonymous-data--m9-.pdf

Office for National Statistics (2013) *Impact of changes in the National Accounts and economic commentary for Q1 2013: http://www.ons.gov.uk/ons/guide-method/method-quality/specific/economy/national-accounts/articles/2011-present/impact-of-changes-in-the-national-accounts-and-economic-commentary-for-2013-quarter-1/impact-of-changes-in-NA.pdf*

Office for National Statistics (2013) *Producer Price Index, Statistical Bulletin* available at: http://www.ons.gov.uk/ons/dcp171778_305483.pdf

Office for National Statistics (2013) *Quarterly National Accounts, Q1 2013 Statistical Bulletin* available at: *http://www.ons.gov.uk/ons/dcp171778_314093.pdf*

Office for National Statistics (2013) *Retail Sales, Statistical Bulletin* available at: *http://www.ons.gov.uk/ons/rel/rsi/retail-sales/february-2013/stb-rsi-february-2013.html#tab-background-notes*

Office for National Statistics (2013) *Updated Analysis: Why is GDP revised?: http://www.ons.gov.uk/ons/guide-method/method-quality/specific/economy/national-accounts/articles/2011-present/why-is-gdp-revised-/updated-analysis--why-is-gdp-revised-.pdf*

Office for National Statistics - *Introduction to the Inter-Departmental Business Register (IDBR):* http://www.ons.gov.uk/ons/about-ons/products-and-services/idbr/index.html

Office for National Statistics - *Opinions and Lifestyle Survey*: http://www.ons.gov.uk/ons/about-ons/products-and-services/opn/index.html