

Guidance for Measuring and Reporting on Quality when Administrative Data is used to Supplement or Replace Survey Data

Office for National Statistics

1. Introduction

Administrative data is increasingly being used to supplement or replace survey data. Because of this it is important to understand what quality issues can arise when using and introducing administrative data, the quality of outputs based on administrative data, and how to report on quality of administrative data.

The purpose of this guidance is to help statistical producers:

- ensure administrative data used at ONS is of the best quality possible
- ensure that there are limited quality issues as a result of the transition from survey to administrative data
- ensure that any changes to quality are reported on effectively
- ensure all quality information that has an impact on the data is reported on effectively
- raise awareness of ways in which the use of administrative data may change the quality of statistical outputs

It is important that we measure and report on the quality of our data and statistics so that users can be confident when using them. These processes also help business areas to comply with Principle 4 of the Code of Practice for Official Statistics “Sound methods and assured quality”, along with Principle 1 “Meeting user needs”.

1.1 Code of Practice

It is important to comply with the UK Statistics Authority [Code of Practice for Official Statistics](#) when introducing a new administrative data source.

Business areas should be careful to ensure that the outputs using new data being introduced comply with the Code of Practice, as does the output, which may have gone through changes in methodology as a result of the introduction of administrative data.

The Code of Practice is important as it shows users that they can be confident in the statistics we produce and can use them as a basis for making decisions.

One important tool that will help statistical producers comply with the Code of Practice is the [Administrative Data Quality Assurance Toolkit](#). The use of this toolkit is a requirement for all areas using administrative data and is the mechanism that the Authority will use to determine compliance with the Code.

1.2 Structure of Guidance

This guidance comes in 2 main parts.

The first part focuses on guidance for measuring quality and quality assuring, both before the administrative data are used and throughout the process.

The second part focuses on reporting on quality. This includes how and where to report quality, and what quality information needs to be communicated to users.

2. Measuring Quality and Quality Assurance

This section of the guidance gives information on what things should be considered when measuring the quality of, and quality assuring administrative data and administrative data based outputs. The guidance is split into sections based on overarching themes, these are; Cost and Efficiency, Technical and Data Considerations, and ESS Quality Dimensions. This was split into these themes as we wanted our guidance to be coherent with the ESS Quality Dimensions where possible. Guidance that fell outside of the scope of these dimensions fitted logically into the other themes.

Business areas should obtain some of these quality measures from those supplying the data. For example, suppliers may have information on item and record completeness that is useful for business areas to understand the quality of the data they are receiving.

Business areas should use this guidance to help determine whether an output is fit for purpose and as a basis to help business areas communicate relevant quality information to the users of the statistics.

2.2 Cost and Efficiency

When introducing an administrative data source to supplement or replace survey data, business areas should consider what effect this may have on cost and efficiency. This includes:

Before the data are used:

- **If administrative data are used to supplement a survey, whether and how much will the survey sample be reduced.**
- **Whether there will be a change in respondent burden (compliance cost) for those filling out surveys if it is supplemented or replaced by administrative data.**
→ In this case we mean whether the use of administrative data will mean that fewer units have to be sampled or the units sampled have to answer fewer questions, and what effect this will have on compliance costs for the survey.
→ For more information on calculating compliance cost changes please contact the Survey Control and Compliance Unit: SCCU@ons.gov.uk
- **Whether there will be any changes in cost to the business area from introducing administrative data in terms of both data collection and the output production process.**
→ For example, there are likely to be additional costs associated with processing a new administrative source to be included in a statistical output.
- **Whether there will be any changes in efficiency for the business from introducing administrative data; will processes be more efficient, or will an increased amount of processing be put in place?**

Throughout the process:

- **Whether the business area will be putting an increased burden on those supplying the**

data (to prepare the data for statistical purposes).

→ For more information on this please consult [this document](#) called “Guidance for Measuring Supplier Burden and Changes in Respondent Burden when using Administrative Data”.

2.2 Technical and Data Considerations

Business areas should consider the technical aspects of receiving, securely storing, and using administrative data. Administrative data are often received as large quantities of data and metadata, and so a number of technical issues should be considered. Things to be considered include:

Before the data are used:

- How much data would have to be received and processed, and, where and how it can be stored.
- How the data are collected.
→ This includes what type of organisation the data are being collected from and whether they are under any authority or legislation with regards to how the data are collected. Also any cleaning they might do before sending the data, along with how compatible this is with the needs of the business area or output.
→ To understand this it would be helpful to liaise with the data supplier in order to understand any potential issues with the data.
- If more than one administrative data source will be used, how they will be linked together; what systems are in place to link data and how likely is there to be linkage error?

Throughout the process:

- Could any of the data received be reused for other purposes?
→ When considering this it is important to remember that the use of data for other purposes would have to be allowed by the legal agreement for accessing the data.
- How readily useable are the files and the data within the files that were received from the supplier?
→ For example, how much pre-processing is required for taking on the files.
- What processes are in place to clean, categorise, and link the data; will this editing and processing suit the needs of the output.
- Whether the administrative data are being used to combine with or supplement survey data.
→ If so, what systems and methods are in place to combine the data, what possible errors could occur in this process (for example linkage error), and how these errors could be mitigated?
- Security: are you able to receive, store, and process the data in a secure manner?

2.3 ESS Quality Dimensions

Business areas introducing a new administrative source should also consider the ESS quality dimensions and what effect introducing a new administrative source will have on these.

Business areas should consider the needs of the users and what their quality requirements are when considering whether data and statistics are of sufficient quality. For example users may need an output to be timely but are not concerned about comparability internationally.

Therefore when measuring the quality of that statistic, there is no need to be overly concerned if the statistics are not comparable internationally but there would be if the output was not timely.

Business areas also need to make sure that the impact on quality due to the introduction of administrative data is understood and the outputs are still of acceptable quality; along with making sure that the benefits of introducing administrative data outweigh any potential costs.

Similarly, business areas should consider whether there will be any trade-offs between quality dimensions as they are not independent of each other. This means whether the introduction of administrative data causes an improvement in quality in one of the dimensions and a worsening in quality in another. Business areas should consider what effect any trade-offs have on overall quality and whether it is a benefit to have less quality in one dimension in order to have an improvement in another dimension. Trade-offs between quality and cost will also need to be considered and appropriately balanced. Any such trade-offs should be clearly explained to the users of the statistic.

Business areas should also remember that quality changes when introducing administrative data are not consistent. Different sources and outputs will have different challenges and there is no “one size fits all” approach in measuring quality changes. However, there are a number of possible issues within each quality dimension that should be considered.

2.3 Quality Considerations within Each ESS Quality Dimension

Possible quality issues within each quality dimension are discussed below.

2.3 Accuracy and Reliability

This dimension considers how close the output is to the true results. Here business areas should consider the following issues:

Before the data are used:

- **Whether data are recorded correctly and appropriately.**
- **How the data owners collect and process the data.**
→ **Particularly what quality or validation checks are carried out by data collectors?**

Throughout the process:

- **Coverage – business areas should consider whether the data has any under- or over-coverage (whether units in the target statistical population that should be counted are not counted, or if units that are not in the population are counted when they should not be).**
→ **It is important to remember that data sources can have both under- and over-coverage.**
→ **An example of both under- and over-coverage is if we wanted to collect administrative data on all businesses in Great Britain but the administrative data source included data from Northern Ireland (over-coverage) and no data from Wales (under-coverage).**
- **Completeness – business areas should consider how complete the administrative dataset is and how this compares to the survey response rate.**

→ Is the level of completeness higher than the survey response rate or not as good is one thing to consider.

→ However, just because completeness is higher than the survey response rate does not mean that it is better quality data, a survey with a low response rate may still be representative of the population and administrative data with a high completeness may be biased.

→ It is important to base the decision to use a data source not just on how complete the data are, but also how useful they are and what trade-offs there may be with other quality dimensions. For example data may become more complete over time but this would have a negative effect on timeliness.

- Linkage errors – if the data needs to be linked are there linkage errors and how can these be mitigated.
- Data errors – are there errors in the data and can these be dealt with.
- Whether there are possible errors in processing made by the data owners.
- Bias and precision: these are concepts that are easily defined for survey estimates, but less easy to define and measure on administrative data based statistics. Nonetheless, it is useful to consider similar ideas when using administrative data: do the administrative data measure what they are supposed to (for administrative purposes); do they measure the statistical requirement accurately; what errors may there be in the statistic because of incompleteness in the administrative dataset.

2.3 Coherence and Comparability

This dimension considers the extent to which different outputs using the same statistical processes use the same concepts and harmonised methods, and the degree to which the data can be compared over time, region, and so on. Here business areas should consider the following issues:

Before the data are used:

- Are definitions in the administrative data source similar to those in the output.
 - For example, if turnover data is being taken from an administrative data source is this the same or similar to the definition of turnover that is used in the output, if not, are there any methods that would make it more coherent?
 - This is an issue that is related to the harmonisation of concepts and definitions in outputs that the ONS Harmonisation team are working to improve. For more information please contact the Harmonisation team: harmonisation@ons.gov.uk
- To what extent methods would need changing and what effect would this have on coherence with previous publications.
- Would there be any discontinuity and how would this affect the use of the data.

Throughout the process:

- Are the administrative data comparable to the previous survey data.
 - For example will the new methods mean there is a break in the time series.
- How does the output using the new source compare to similar outputs internationally.
- If geography codes are used, ensure they are consistent with those recommended in the GSS Geography policy.
 - For more information, see <https://data.gov.uk/dataset/gss-geography-policy1>

2.3 Timeliness and Punctuality

This dimension considers the time between data production and publication of the output, and time between actual and planned date of publication. Here business areas should consider the following issues:

Before the data are used:

- How long it takes for data owners to collect the data.
→ This may be a trade-off with completeness.
- Could there be delays in updating administrative databases and how this would affect the output.
- How long it will take to receive the data after it is collected by the data owner – whether there are any possible delays.
- Would there be a change in the lag between the reference period and the release of estimates.
- Is there a difference between time periods for which data are collected and time periods for which it is needed.
→ For example are the data collected on a quarterly basis but needed for monthly estimates.
- Whether additional time is required by the business area to process and use the administrative data in their outputs and how this may impact on the timeliness of releases.

→ One way of dealing with some of the potential issues above would be to consider producing provisional estimates and then releasing revisions to the output. However, statistical producers would need to consider the practicalities of this and trade-offs with other dimensions. If revisions were used then the reasoning behind using them should be communicated to the users of the statistics so that they understand how this change would affect the use of the statistic.

2.3 Accessibility and Clarity

Whilst most issues related to accessibility and clarity are general issues that should be considered regardless of whether the statistics are administrative or survey data based, there are some issues that are specifically related to administrative data that should be considered.

→ For more information on the general issues to consider for “Accessibility and Clarity”, please consult the [European Statistics Code of Practice](#)

Before the data are used:

- Whether the data owner will allow users of statistics have access to the microdata as part of the legal agreement between the data owner and the business area.

Throughout the process:

- Whether the metadata and accompanying information that is received from the data supplier is in enough detail for users to understand how the data are put together and processed.

2.3 Relevance

Whilst most issues related to relevance are general issues that should be considered regardless of whether the statistics are administrative or survey data based, there are some issues that are specifically related to administrative data that should be considered.

→ For more information on the general issues to consider for “Relevance”, please consult the [European Statistics Code of Practice](#)

Before the data are used:

- Whether the administrative data meets the needs of the users as well as the previous survey data did, or if it is better at meeting user requirements.
→ For example, if you get survey data from the whole of the UK for a statistic but the administrative data only covers Great Britain would that data still meet the needs of the user, considering Northern Ireland is no longer covered?

Throughout the process:

- Whether any issues with the data or output that affect any of the other quality dimensions also affect the relevance of the output.
→ For example, if there are any issues with the timeliness of the output then this could have the potential to make the output less relevant.

2.4 Useful Tools

[Administrative Data Quality Assurance Toolkit](#)

Many of the potential quality issues with administrative data are related to how the data are collected, stored, quality assured, and delivered by the data collectors, owners, and suppliers. Because of this it is important that business areas keep in contact with suppliers and owners and make sure they understand the importance of these processes.

The Office for Statistics Regulation has developed a toolkit that business areas are expected to use. This gives information on working with those outside of Statistical Institutions so that we receive the best quality of data possible. It also gives guidance on how much quality assurance an administrative dataset needs based on the level of risk of quality concerns and on the public interest. The Office for Statistics Regulation website also contains guidance information to help statistical producers use the toolkit.

[Checklist for the Quality evaluation of Administrative Data Sources](#)

Another useful tool is one developed by Statistics Netherlands, the “Checklist for the Quality evaluation of Administrative Data Sources”. This checklist assists quality evaluation of source and metadata aspects of administrative data and is used by many different NSIs across Europe.

[ESSnet: Use of Administrative and Accounts Data in Business Statistics](#)

Many of the areas to consider that have been mentioned in this guidance relate to the quality indicators set out in the ESSnet on “Use of Administrative and Accounts Data in Business Statistics” Workpackage 6: “Quality Indicators when using Administrative Data in Statistical Outputs”. The document linked gives in depth explanations of how to calculate the quality of each indicator within each ESS quality dimension. It is a very useful and comprehensive tool that will help business areas gain an in-depth understanding of the quality of their output.

3. Reporting on Quality

Based on the guidelines in the first section of this report, business areas should report on the quality of the new statistic and any changes in quality from the old output to the new output. They should report on any changes in quality that are the result of introducing administrative data to supplement or replace survey data, and the resulting methodological changes. They should also report on how these impact on the use of the data.

Of particular importance when using administrative data is to list the sources that are used to create the output as it is important that users understand where the data come from. However, it may not be possible to list all of the sources used, if this is the case then the main sources should be stated and an overview should be given.

3.1 Quality and Methodology Information (QMI)

ONS outputs have associated QMI reports that contain static quality information about an output. It is designed to inform users about the strengths and limitations and what the data is suitable to be used for.

If an output is introducing an admin data source then it needs to change its QMI report to reflect the new quality and methodology information.

Here is the link to further information about the QMI report along with a template of the report; this contains all the information needed to fill out a QMI report for non-survey data.

3.2 Quality Summary Page

Quality Summary Pages are also a useful way of communicating quality information to users in an easily accessible way. The aim of a Quality Summary Page is to give users the information they need so as not to misuse the data.

[Here](#) is an example of Quality Summary Page. Quality information is also displayed on the main statistical bulletin for the output and is the information displayed in the section “Things you need to know”. For more information please contact [Sarah Tucker](#) at the Quality Centre.

3.3 Quality Information within Statistical Bulletins

Statistical Bulletins should contain quality information that helps users understand data and quality implications for the data in specific releases.

Daisy Hamer (daisy.hamer@ons.gov.uk)

The Quality Information provided should help users understand how to use the data reported on in the Statistical Bulletin.

4. Annexe

[UK Statistics Authority Code of Practice for Official Statistics](#)

[Administrative Data Quality Assurance Toolkit](#)

[Guidance for Measuring Supplier Burden](#)

[European Statistics Code of Practice](#)

[Checklist for the Quality Evaluation of Administrative Data Sources](#)

[ESSnet: Use of Administrative and Accounts Data in Business Statistics](#)

For more information please contact:

Daisy.hamer@ons.gov.uk

Or

Ons.qualitycentre@ons.gov.uk