



Quality Assurance of Administrative Data

Setting the Standard

Version 1 January 2015

On 31 July 2014 the UK Statistics Authority published an Exposure Draft report [Quality Assurance and Audit Arrangements for Administrative Data](#).

This report presents the Authority's regulatory standard that it will use to assess statistics derived from administrative sources against the *Code of Practice for Official Statistics*. It is published alongside the [Administrative Data Quality Assurance Toolkit](#).

The Authority will release further supporting information to guide both statistical producers and non-statistical officials in using administrative data for official statistics:

- Emerging case studies of good practice
- Guidance for policymakers when using statistics
- Response to consultation

Introduction

1. This regulatory standard confirms the quality assurance arrangements that are required for statistics that are compiled using administrative data to comply with the *Code of Practice for Official Statistics*¹. The *Administrative Data Quality Assurance Toolkit* (QA Toolkit) is the mechanism that the Authority will use to determine compliance. The Toolkit is built around the Quality Assurance Matrix (QA Matrix) which presents the levels of assurance for four areas of practice related to the quality assurance of official statistics and of the administrative data used to produce them.
2. Our assessors will:
 - Provisionally determine their view of the appropriate level of assurance required for a set of statistics.
 - Request from statistical producers:
 - a) an explanation of their judgments of the required level of assurance for a set of statistics;
 - b) evidence to support the rationale for these decisions;
 - c) evidence of the actions, and the rationale for deciding upon those actions, that the producer has taken to achieve compliance with the chosen level of assurance; and
 - d) evidence which demonstrates that the statistical producer has embedded practices for keeping its quality assurance arrangements under review.
 - Identify whether the submitted evidence complies with the *Code of Practice*.
 - For instances of non-compliance, require statistical producers to take action to meet the standard of the *Code of Practice*.
3. This paper sets out the basis of the QA Toolkit in relation to the *Code of Practice*. It also addresses a number of the issues raised by respondents to the Exposure Draft, in relation to the definition and nature of administrative data, the need for repeated and

¹ <http://www.statisticsauthority.gov.uk/assessment/code-of-practice/index.html>

ongoing evaluation of data quality, the complex nature of data supply relationships, and the role of systematic external audit.

4. The QA Toolkit recognises that a proportionate approach is appropriate, whereby some statistics will need greater levels of assurance than others. The levels can be determined by an evaluation of: (i) the likelihood of quality concerns arising in the administrative data that may affect the quality of the statistics; and (ii) the nature of the public interest served by the statistics.
5. The *Code of Practice* requires statistical producers to provide explanatory information to help users understand the strengths and limitations of official statistics in relation to their use. Producers need to ensure that they have a good appreciation of the impact of the data collection processes on the statistics and that they communicate these issues to those who use the statistics. It is important to critique the data at each stage of the statistical production process.
6. The Authority encourages statistical producers to use this Toolkit routinely as part of their analysis and monitoring of administrative data systems, and to share their findings with users.

Background

7. On 31 July 2014 the Authority published an Exposure Draft report *Quality Assurance and Audit Arrangements for Administrative Data*. The Exposure Draft presented the Authority's proposed regulatory standard for the quality assurance of administrative data that are used to produce official statistics. We considered the risks associated with the use of administrative data for statistical purposes, identified some examples of existing practices across government in addressing those risks, and presented our proposed mechanisms for statisticians to use when seeking to implement best practice.

Consultation

8. The Authority invited comments on the proposed approach. In response we received 14 written comments from private individuals, government departments and from statisticians responding in personal and professional capacities. Between September and December 2014 we spoke directly to over one hundred statisticians in government departments across the UK, as well as to some policymakers, public and private sector data suppliers and other regulatory bodies.
9. Overall, the feedback endorsed the proposals and reassured us that the approach was sound. It was clear that there is a common understanding that it is important for statistical producers to gain, and share with users, a fuller understanding of the administrative data that they use to produce official statistics, of the circumstances in which they are produced, and how they are tested and verified. Statisticians told us that they face many challenges when using administrative data, including having limited resources and obtaining data from old, poorly documented data systems. Statisticians also told us of some frustrations in having to react to policy changes rather than being consulted as part of the process. There was also a request for examples of good practice

to be shared between statistical producers. We also heard from dynamic and enthused statisticians determined to implement improved processes and who shared with us good examples of their work.

10. We will publish supporting documents on our website, these will include:
- revised *Guidance for policymakers using statistics*;
 - a series of case studies that demonstrate good practice in aspects of quality; assurance
 - key responses to the consultation and the Authority's response.

The Code of Practice for Official Statistics

11. The *Code of Practice* encompasses the production, management and dissemination of official statistics – this covers the entire statistical process from the identification of needs, to the decision to collect or compile data, through to providing advice to the user. It requires statistical producers to ensure that administrative sources are fully exploited for statistical purposes, with appropriate safeguards in place² and that they:

- Use data based on definitions and concepts which approximate well with those required for the statistics³.
- Inform users about the quality of their statistical outputs, including estimates of the main sources of bias and other errors in the data⁴ and the quality and reliability of statistics in relation to the range of potential uses⁵.
- Publish supporting metadata which includes information about the quality assurance procedures⁶, quality guidelines⁷, the administrative systems used in the production of official statistics⁸ and those with the potential to be so used⁹, and the arrangements for auditing the quality of the data¹⁰.
- Identify arrangements for allowing statistical staff access to administrative data for statistical purposes and ensure that full account is taken of the implication for official statistics when changes to administrative systems are considered¹¹.
- Promote statistical purposes actively in the design of administrative systems in order to enhance the statistical potential of administrative records¹².
- Maximise opportunities for the use of administrative data, cross-analysis of sources and for the exchange and re-use of data, to avoid duplicating requests for information¹³.

² Protocol 3 of the *Code of Practice for Official Statistics*

³ Protocol 3, Practice 2 of the *Code of Practice for Official Statistics*

⁴ Principle 4, Practice 2 and Protocol 1, Practice 4 of the *Code of Practice for Official Statistics*

⁵ Principle 8, Practice 1 of the *Code of Practice for Official Statistics*

⁶ Principle 4, Practice 3 of the *Code of Practice for Official Statistics*

⁷ Principle 4, Practice 4 of the *Code of Practice for Official Statistics*

⁸ Protocol 3, Practice 5(a) of the *Code of Practice for Official Statistics*

⁹ Protocol 3, Practice 5(c) of the *Code of Practice for Official Statistics*

¹⁰ Protocol 3, Practice 5(e) of the *Code of Practice for Official Statistics*

¹¹ Protocol 3, Practice 5(b) and 5(d) of the *Code of Practice for Official Statistics*

¹² Principle 6, Practice 3 of the *Code of Practice for Official Statistics*

¹³ Protocol 3, Practice 3 of the *Code of Practice for Official Statistics*

Administrative data

12. Administrative data refers to information collected primarily for administrative reasons (not research). This type of data is collected by government departments and other organisations for registration, transactions and record-keeping, usually when delivering a service¹⁴. Administrative data are often used for operational purposes and their statistical use is secondary.
13. The Exposure Draft outlined some of the benefits and the challenges of using administrative data to produce official statistics. The purpose of this paper is to introduce the regulatory standard rather than set out a definitive description of those merits and risks.
14. The Authority recognises that there are limitations with administrative data and that these can create complications when compiling official statistics. However their use is central to the production of official statistics and the existence of these challenges places a premium on proactive quality assurance to investigate the data, manage identified issues, and clearly communicate any limitations to users.

Quality Assurance of administrative data

15. Quality assurance of administrative data is more than simply checking that the figures add up. It is an ongoing, iterative process to assess the data's fitness to serve their purpose. It covers the entire statistical production process and involves monitoring data quality over time and reporting on variations in that quality. Post-collection quality assurance methods, such as data validation, are an important part of the quality assurance process, but can be of limited value if the underlying data are of poor quality. The Authority encourages the application of critical judgment of the underlying data from administrative systems *before* the data are extracted for supply into the statistical production process. As with survey data, producers need to: investigate the administrative data to identify errors, uncertainty and potential bias in the data; make efforts to understand why these errors occur and to manage or, if possible, eliminate them; and communicate to users how these could affect the statistics and their use.

Using the Quality Assurance Matrix

Practice Areas

16. In an Assessment of official statistics based on administrative data, the Authority will consider the quality assurance of the data across the four practice areas outlined in the QA Toolkit. These practice areas demonstrate the need for the quality assurance of statistics obtained from administrative statistics to extend beyond the checks made by statistical producers on the data they receive. The producers should demonstrate knowledge of the operational context in which the data are recorded, and an understanding of the impact that the motivations of data inputters can have on the data.

¹⁴ Definition from the Administrative Data Research Network <http://www.adrn.ac.uk/faq/about-the-data>

They should also have good communication links with data supply partners and understand their partners' data quality processes and standards.

Four practice areas associated with data quality

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

The Quality Assurance Matrix

- The Authority recognises the resource challenges faced by statistical producers and supports a proportionate and pragmatic approach to the assurance of administrative data in relation to each area of practice. The QA Matrix encourages producers to consider the level of data quality issues that could affect their data together with the level of public interest in the statistics.
- The QA Matrix is **not** intended to be used as a check-list or a tick box exercise. Work has been done to develop such lists¹⁵ and they can be useful in developing quality assurance processes. However, over-reliance on a checklist may lead to cases where issues are over-simplified and limit the ongoing development of quality assurance processes. The QA Matrix presents examples of the types of evidence that assessors would expect to see as part of an Assessment, to provide statistical producers with an understanding of the standard required for each level of assurance.
- The assurance levels set out in the QA Matrix are not intended to be interpreted as a RAG status¹⁶ as commonly used in project management. For example level A3 does not necessarily mean that remedial action is needed; instead it could mean that the statistics have ongoing levels of higher public interest and higher quality concerns which would mean that higher levels of assurance would be continual. Remaining at A3 level of

¹⁵For example see <http://essnet.admindata.eu/WorkPackage?objectId=4257>

¹⁶ RAG (Red, Amber, Green) status is commonly used in project management to visually highlight the risk level of a project, with red showing the greatest risk and green the least.

assurance over time does not represent a failing by the statisticians; it simply demonstrates that a higher level of assurance is necessary for a certain set of statistics.

20. The QA Matrix is not exhaustive. It does not present all the possible types of work that statistical producers could carry out to assure the quality of their data. In many cases they will be undertaking other types of quality assurance work which they should share with assessors. Similarly, statistical producers might recognise that not all elements suggested at a certain level of assurance will be appropriate for their statistics, and have appropriate reasons for making such a judgment. Ultimately it is for each statistical producer to decide how it meets the standard and present this evidence to the Authority.
21. A key element of practice emphasised in the QA Toolkit is that if flaws are found in administrative data, statistical producers should: evaluate the likely impact on the statistics; establish whether the issue can be resolved, or whether there is any other action they can take to mitigate the risks; and determine whether the level of impact is such that users should be notified. It is recognised that often issues discovered through quality assurance are complex and will require time and staffing and financial resources to address. Statistical producers are required to maintain ongoing compliance with the *Code of Practice*. If, in the course of these investigations, a statistical producer discovers a systemic issue in the administrative data that has a substantial adverse impact on the statistics, we encourage the statistical Head of Profession to contact the Authority to discuss appropriate action.
22. The Authority recognises that there are certain circumstances in which regular, systematic external evaluation, audit or inspection of the underlying data is essential to increase both the quality of, and public confidence in, statistics produced from administrative data. For statistics requiring higher levels of assurance, these external evaluations should be regular and repeated. In the absence of regular repeated external scrutiny (for those sets of statistics for which it is appropriate) the statisticians should highlight the deficiency for users and investigate whether there are other sources that could be used.
23. Administrative data underpinning official statistics can be subject to, or feature in, various kinds of audit, depending on their operational context, for example: financial, clinical, and statistical audit in which a sample of existing cases is investigated. These investigations can be used to provide context, and in some cases corroboration of the data quality. It is good practice to investigate whether these other types of audits could provide information to support the statisticians' quality judgments, and to use them if appropriate. It is also recognised that in some cases these types of corroborating information are not available.
24. The steps to be taken by statistical producers need to go beyond a narrow interpretation of 'quality assurance'; they should also encompass the working arrangements and relationships with other agents, particularly data supply partners. These working arrangements can range from straightforward data transfers from a single data supplier, through to more complex large-scale systems. In addition some official statistics are based on administrative data provided by another statistical producer body.

For statistics based on administrative data obtained from outside the Government Statistical Service some examples include:

- **Directly from the organisation that records the data** – in this case the statistical producer should engage with this organisation directly and take into consideration the detail and nature of information received from the data supplier when deciding on appropriate quality assurance. The responsibility for producing information about the quality of the statistics lies with the producer. In the case of large numbers of direct suppliers, the statistical producer might explore other ways of engaging, such as holding meetings to discuss common quality issues and liaising with information governance groups.
- **Via intermediary organisations** – for example local authorities collect data and pass these to the producer – in this case the statistical producer should: 1) have an understanding of the entire data cycle and the quality assurance processes carried out by the original data supplier; and 2) engage with intermediary organisations to understand their quality assurance processes and standards.

For statistics based on administrative data collected by another government department or official statistics producer body:

- In this case both the statistical producer and the supplier department have a responsibility to understand and communicate about the quality of the data. The producer with responsibility for publishing the statistics has the ultimate responsibility for assuring itself about the quality of the underlying data and communicating this to users. However the statistical Head of Profession for the supplier partner body has a responsibility to ensure that it shares its quality assurance information with the statistical producers, including clear information about any limitations or bias in the data, and takes action to address any queries or concerns raised by the statistical producer. In cases where producer bodies share data for statistical purposes, or both work with the same data supply partners, we encourage such producers to work together to develop a better understanding of the quality of the administrative data, sharing intelligence and insight about the data with each other and with users.

Guide for policymakers

25. As part of the suite of reports published alongside this paper, the Authority will publish *Guidance for policymakers using statistics*. This will build upon Annex A of the Exposure Draft report. As part of the feedback we received to the Exposure Draft, statisticians told us of the need to share with policymakers the benefits and limitations of official statistics when policy decisions are made. Statistical input is an important part of policy making and this guidance will point to critical questions that policymakers can ask about the statistics they use and of the statisticians who produce them.