

The assurance of administrative data: A proportionate approach¹

Penny Babb

UK Statistics Authority, London, 1 Drummond Gate, London, UK

Tel.: +44 20 7592 8666; E-mail: penny.babb@statistics.gov.uk

Abstract. Administrative data provide an attractive means of obtaining generally cheap and comprehensive data for producing official statistics. However, their limitations need to be understood to ensure the appropriate use of official statistics. The UK Statistics Authority developed a pragmatic and proportionate approach to the quality assurance of administrative data, with three levels of assurance: basic, enhanced and comprehensive. The level selected is dependent on the risk of data quality concerns and degree of public good represented by the statistics. The approach identifies four practice areas that reflect the need for the assurance of administrative data to extend beyond the checks made by statistical producers on the data they receive. Producers should also have a good understanding of how the data are compiled and why; have established effective communication links with data supply partners; and understand their partners' data quality processes and standards. The Authority's guidance is being applied by UK statistical producers, to achieve compliance with the *Code of Practice for Official Statistics*.

Keywords: Administrative data, quality, regulatory standard

1. Introduction

1.1. Administrative data and official statistics

Administrative (or 'admin') data sources are those in which data are collected for an operational or administrative purpose, with statistical use being a secondary purpose. Such data sources have been central in the production of official statistics in the UK for over 150 years. While sample surveys have been subject to intense methodological and statistical investigation, to develop techniques that overcome some of their inherent limitations, that has not been the case for admin data. Instead, official statisticians can be more trusting, or may be more resigned to the weaknesses, of these data sources. Within the UK, and perhaps other countries too, this has resulted in limited insight into the potential limitations and biases affecting official statistics based on admin data.

1.2. Police recorded crime

A stark illustration of the loss of confidence in official statistics that can accompany data quality problems occurred recently in England, when a serving police officer raised concerns within a UK parliamentary committee hearing [1] about the integrity of data recording by police forces. The whistleblower reported the mishandling of recorded crime records within the Metropolitan Police Force – the police service for London. These claims held serious implications for the accuracy of the official statistics. Her Majesty's Inspectorate of Constabulary (HMIC), the regulator for the police service, then substantiated the scale of the issue, identifying that around 20% of crimes may be under-recorded across the police forces in England [2]. Over 800,000 crimes reported to the police may have gone unrecorded each year. The UK Statistics Authority [3] removed the National Statistics designation of the police recorded crime statistics and required significant improvements to the statistics, including the assurance of the data.

¹This paper is based on a presentation by the author at the Q2016 Conference on Quality in Official Statistics.

1.3. The broader issue

More generally, the problems highlighted concerns about how well official statisticians understand the admin data sources they use, and whether they are sufficiently aware of quality issues that affect their statistics. As a result, the regulatory team in the Authority identified the quality assurance (QA) approaches used by statistical producers. The Authority published a guidance document, *Quality Assurance of Administrative Data* (QAAD) [4]. Central to this guidance is the QA Matrix, based on four pillars of statistical practice.

Of course, statistical producers use many admin data sources in producing the wide range of official statistics published today. Few sources perhaps match police recorded crime for the scale of concern around the integrity of the data, those being particularly susceptible to the distortive effects of performance targets. And so the Authority recognises the need for assurance that is pragmatic and proportionate. As a result, the QA Matrix has three levels of assurance: basic, enhanced and comprehensive.

The Authority expects statistical producers in the UK to apply this guidance in quality assuring the admin data they use to produce official statistics. The Authority applies its regulatory standard [5] in judging how well producers understand and explain the strengths and limitations of their statistics, including the assurance of the admin data, when assessing compliance with the *Code of Practice of Official Statistics* [6].

2. Background

2.1. UKSA – its role and responsibility

The UK Statistics Authority was established in April 2008 under the *Statistics and Registration Service Act 2007* [7] and operates independently from government. It has a dual responsibility for the oversight of the production of official statistics by the Office for National Statistics and for the regulation of official statistics across the UK. The Authority is responsible for the *Code of Practice for Official Statistics*. All official statistics should comply with the *Code*. The Authority determines the extent to which official statistics meet the *Code* through its process of Assessment [8]. Official statistics achieving the highest standards of trustworthiness, quality and value are awarded the National Statistics designation. Only the Authority can award and remove the designation.

2.2. The development of the regulatory standard

The Authority's focus on the quality of admin data centred on the development of a regulatory standard to apply when assessing the compliance of official statistics based on admin data against the *Code*. The QAAD guidance which accompanies the regulatory standard describes the statistical practices for quality assuring admin data [9].

The QAAD guidance was developed from the feedback obtained in many in-depth interviews with more than 100 statisticians. These conversations built a picture of the assurance steps taken and the issues encountered. The guidance was also informed by tools developed for, and by, national statistical institutes, to understand the best international practices for assuring admin data, including those from Statistics Netherlands and the US Census Bureau. The Authority's conceptual approach to assuring admin data was summarised in *Quality Assurance and Audit Arrangements of Administrative Data – Exposure Draft* [10].

3. Quality assurance of administrative data (QAAD)

3.1. QAAD framework

Quality assurance of admin 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 data are of poor quality. It is not enough for statistical producers to quality assure the data received from data suppliers but a critical judgment should be made of the quality of data from admin systems *before* extraction and supply into the statistical production process. To do this, information is required on the data collection and entry, processing, validation and assurance by data suppliers.

The four QAAD pillars of statistical practice are: operational context and admin data collection; communication with data suppliers; suppliers' quality assurance principles, standards and quality checks; and, the producer's quality assurance investigations and documentation. These practice areas demonstrate the need for the quality assurance of official statistics to extend beyond the checks made by statistical producers on the data they receive.

3.1.1. *Operational context and administrative data collection*

Operational context reflects the need for statistical producers to gain an understanding of the environment and processes producing the admin data. They need to look for the factors which might increase the risks to the quality of the data – such as the effects of targets and performance management regimes, the numbers of data collector and supplier bodies, and the information governance arrangements. The use of targets and performance management regimes may affect the recording of data, particularly if the target definitions are ambiguous or complex, or there is scope for different interpretations and practices within the operational bodies – for example, in health service waiting times, the approach taken to starting and stopping the clock in relation to treatment may vary between hospital trusts. The ways in which these risks are mitigated (i.e. the safeguards) should be identified and their effectiveness evaluated. Preparing a process map can help statistical producers identify the risks and design safeguards.

The admin data collection process should be described, identifying, for example: the definitions, classifications and codes used in recording the data; any variations across data suppliers; and the nature of data collected – such as whether all items are objective or also include subjective information. It is common to think of data collected in admin systems as simple and homogeneous, the result of routine processes. However, ‘data’ is a term referring to a collection of information whose nature can vary widely. Objective data items include transactional information, such as, whether a payment has been made, or event-recording such as the notification of death. In contrast, subjective data items, such as a person’s ethnicity or occupation, rely on information that can only be provided by a respondent and cannot be verified by the system itself. Internal validity checks can only be used to confirm that the code used is consistent with the permitted coding rules; they cannot check the accuracy of the information recorded.

3.1.2. *Communication with data suppliers*

Communication with data suppliers is vital. Effective relationships with suppliers should be based on detailed written agreements (such as in a service level agreement or memoranda of understanding), including change management processes, to ensure that statistical needs are considered when changes are being made to the admin systems and documented data supply arrangements. When multiple data suppliers are in-

involved, producers should ensure that they have a good understanding of the approaches adopted across the sector to ensure consistency in recording and quality levels.

3.1.3. *Suppliers’ quality assurance principles, standards and quality checks*

Statistical producers should understand the validation checks that are conducted by the supplier, and the results of the checks. Some operational systems will also have a process of audit established; for example, checks into financial payments – in which case the scope of the audit and the outcomes should be identified. A supplier may have established its own quality assurance plans or guidelines to determine what it regards as acceptable data quality. It may also have undertaken actions to address weaknesses and conducted or commissioned investigations to assess compliance with quality standards. Producers should identify any steps taken to determine the *accuracy* of the admin data, that is, the closeness of computations or estimates to the true values, as well as their validity.

3.1.4. *The producer’s quality assurance investigations and documentation*

Statistical producers conduct their own quality assurance. These checks should consider whether the derived aggregated statistics are meaningful, and whether changes in trends and discontinuities can be explained – these should include any changes in target definitions and their implications for the statistics. Since their checks cannot, by themselves, verify the accuracy of the admin data, producers should seek additional information. They should corroborate their quality assurance findings against data from other sources, such as surveys or other admin data sources, and compare rates or proportions with the other data sets. And statistical producers should review any investigations undertaken by, or on behalf of, external bodies such as regulators and auditors.

3.2. *Pragmatic and proportional – levels of assurance*

Some statistics require a more thorough and detailed approach to assurance than others. The QA Matrix (see Table 1) helps the Authority’s assessors, as well as statistical producers, determine the types of assurance and documentation required to inform users about the quality of the admin data and any implications for the official statistics. It has three levels of assurance for ensur-

Table 1
QA matrix

Areas of practice related to quality assurance of administrative data regularly provided for producing official statistics – consider the following types of activities:		QA principles, standards and checks applied by data suppliers	Producer's QA investigations & documentation
Level of assurance	Operational context & administrative data collection	Communication with data supply partners	
A1: Basic assurance	<ul style="list-style-type: none"> – Producer has provided users with an outline of the administrative data collection process, – Outlined the operational context, Identified actions taken to minimise risks to quality. 	<ul style="list-style-type: none"> – Producer has outlined the data provision arrangements, – Fed back identified errors to data suppliers and recorded their response, – Sought the views of statistics users about the data and resolved any quality issues reported. 	<ul style="list-style-type: none"> – Producer has established regular QA checks on the received admin data, – Identified the strengths and limitations of the admin data, – Explained the likely degree of risk to the quality of the admin data.
A2: Enhanced assurance	<ul style="list-style-type: none"> – Producer has provided users with a fuller description of the operational context and administrative data collection arrangements, – Identified and summarised potential sources of bias and error in administrative system. 	<ul style="list-style-type: none"> – Producer has agreed and documented data requirements for statistical purposes, legal basis for data supply, data transfer, data protection, – Established an effective mode of communication with data collector and supplier bodies, IT systems, operational/policy officials, – Sought the views/experiences of statistics users and resolved any quality issues reported. 	<ul style="list-style-type: none"> – Producer has provided a fuller description of its own QA checks on the admin data, – Detailed the general approach and findings for specific quality indicators, – Identified the strengths and limitations of the admin data, – Explained the likely degree of risk to the quality of the admin data.
Comprehensive assurance	<ul style="list-style-type: none"> – Producer has provided users with a detailed description of the admin system and operational context, – Identified issues in design and definition of performance measurements and targets, – Identified and described potential sources of bias and error in the administrative system, – Identified and explained any safeguards used to minimise the risks to data quality. 	<ul style="list-style-type: none"> – Producer has established/maintained collaborative relationships, – Has a written agreement specifying roles and responsibilities, legal basis for data supply, data supply and transfer process, security and confidentiality protection, schedule for data provision, content specification, – Regularly communicated with the data collector and supplier bodies, IT systems, operational/policy officials, – Regularly engaged statistics users, resolved any reported quality issues, and held user group conferences. 	<ul style="list-style-type: none"> – Producer has provided a detailed description of its own QA checks on the admin data (including validation, sense and consistency checks), – Given quantitative (and where appropriate qualitative) metrics for specific quality indicators, – Undertaken comparisons with other relevant data sources, – Identified possible distortive effects of performance measurements and targets, – Identified the strengths and limitations of the data and any constraints on use.
Statistical producer has investigated the administrative data QA arrangements, identified the results of independent audit, and published detailed documentation about the assurance and audit		<ul style="list-style-type: none"> – Reviewed quality reports for the received data, – Identified and documented the findings of investigations and audits conducted on the admin data and associated targets (such as internal and operational audits, and external audits by regulators and professional bodies), – Described the implications for the statistics and determined whether the data continue to be satisfactory for official statistics purposes. 	

ing the suitability of the data quality for each of the four practice areas.

The need for investigation and documentation increases at each level of assurance: basic – the statistical producer has reviewed and published a summary of the admin data QA arrangements; enhanced – the statistical producer has evaluated the admin data QA arrangements and published a fuller description of the assurance; comprehensive – the statistical producer has investigated the admin data QA arrangements, identified the results of independent audit, and published detailed documentation about the assurance and audit.

Judgments about the quality of the data for use in official statistics can be pragmatic and proportionate, made in the light of an evaluation of the degree of concern about the quality of the data and the public interest profile of the statistics. The higher the degree of quality concern and public interest, the higher the level of assurance that is required.

3.3. *Risk of data quality concerns*

Data quality concerns may be magnified when there is a greater likelihood of error occurring in the recording of data and of increased difficulties in identifying inaccuracies. For example, when there are many data collector bodies, such as schools or hospital trusts, there is an increased risk of differing local practices – these can lead to inconsistent definitions and codes being used to measure the same concept. The use of targets and performance management regimes can also lead to a distortive effect on the data – whether through deliberate actions, to improve the apparent performance of the organisation, or indirectly, as a result of the local interpretation of target definitions. Concerns about data quality will be lower for a well-defined system with built-in data entry and validation checks, few data suppliers and well-established arrangements for internal audit of the data.

3.4. *The public interest profile*

The public interest profile reflects the importance of the decisions informed by the official statistics. Higher public interest (or value) will occur, for example, where the use of the statistics is required by legislation or informs resource planning and allocation by government or businesses. A lower public interest may arise where the statistics have a narrower relevance and attract little public debate.

4. Conclusion

Applying the QAAD guidance provides statistical producers with a holistic view of quality assurance associated with admin data. The findings from the producer's own quality assurance checks should be supplemented by knowledge gained through reviewing the other practice areas, to inform a published statement that sets out for users the producer's judgment about the quality of the admin data. It is essential that the statistical producer identifies and explains the implications of substantive quality issues of the statistics to support their appropriate use.

Often issues discovered through quality assurance are complex and will require time and staffing and financial resources to address. This review of admin data is not a one-off event, but is rather a process that requires repeated evaluation to understand the implications of changes and allow for the ongoing monitoring of the data quality. The Authority encourages statistical producers to use the QAAD guidance routinely as part of their analysis and monitoring of admin data systems, and to share their findings with users.

Acknowledgments

With thanks to Emily Gleeson, Sandy Stewart and Ed Humpherson who also worked on the development of QAAD.

References

- [1] House of Commons Public Administration Select Committee (PASC). Caught Redhanded: Why we can't rely on Police Recorded Crime. Thirteenth Report of Session 2013-14. London: The Stationery Office; (2014). Available from: <http://www.parliament.uk/business/committees/committees-a-z/commons-select/public-administration-select-committee/news/crime-stats-substantive/>.
- [2] Her Majesty's Inspectorate of Constabulary. Crime-recording: making the victim count. London: HMIC; (2014). Available from: <https://www.justiceinspectorates.gov.uk/hmic/publication/crime-recording-making-the-victim-count/>.
- [3] UK Statistics Authority. Assessment of Statistics on Crime in England Wales. Assessment number 268. London: UK Statistics Authority; (2014). Available from: <https://www.statisticsauthority.gov.uk/publication/statistics-on-crime-in-england-and-wales/>.
- [4] UK Statistics Authority. Quality Assurance of Administrative Data. London: UK Statistics Authority; (2015). Available from: <https://www.statisticsauthority.gov.uk/monitoring-and-assessment/monitoring/administrative-data-and-official-statistics/>.

- [5] UK Statistics Authority. Quality Assurance of Administrative Data – Setting the Standard. London: UK Statistics Authority; (2015). Available from: https://www.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-settingthetstandar_tcm97-44370.pdf.
- [6] UK Statistics Authority. Code of Practice for Official Statistics. London: UK Statistics Authority; (2009). Available from: <https://www.statisticsauthority.gov.uk/monitoring-and-assessment/code-of-practice/>.
- [7] Cabinet Office. Statistics and Registration Service Act 2007, London: The Stationery Office; (2007). Available from: <http://www.legislation.gov.uk/ukpga/2007/18/contents>.
- [8] UK Statistics Authority. Guidance about Assessment. London: UK Statistics Authority; (2016). Available from: <https://www.statisticsauthority.gov.uk/monitoring-and-assessment/assessment/guidance-about-assessment/>.
- [9] UK Statistics Authority. Quality Assurance of Administrative Data – QA Toolkit. London: UK Statistics Authority; (2015). Available from: https://www.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-qualityassurancetoolki_tcm97-44368.pdf.
- [10] UK Statistics Authority. Quality Assurance and Audit Arrangements of Administrative Data, Exposure Draft. London: UK Statistics Authority; (2014). Available from: <https://www.statisticsauthority.gov.uk/monitoring-and-assessment/monitoring/administrative-data-and-official-statistics/quality-assurance-and-audit-arrangements-for-administrative-data/>.