

## Foreword from the Head of Assessment

Administrative data have been widely used to produce official statistics for many years; for some, such as death registrations, for more than 150 years. New technologies are now enabling the greater use of administrative data by users across all sectors. As well as providing new opportunities and benefits, the use of administrative data in official statistics by statistical producers brings with it a responsibility for ensuring that the data are sufficiently robust for these purposes. Equally, it is essential that any strengths and weaknesses are well understood and explained to users.

In the wake of the Authority's decision to de-designate as National Statistics police recorded crime statistics in England and Wales, we launched a programme of monitoring work on the use of administrative data in creating official statistics. Our programme consisted of a review of performance targets and official statistics (to be published separately) and this review, which focuses on the quality assurance and audit of administrative data.

Since 2008, the Authority has used the tool of Monitoring Reviews to supplement our assessments of individual statistics against the *Code of Practice for Official Statistics* (the *Code*). Monitoring Reviews allow us to address systemic challenges or opportunities that affect whole classes of statistics (for example, in a particular sector like health) or the entire population of official statistics (for example, on the use made of official statistics). These reviews have typically examined evidence, brought out systemically important findings, and made recommendations for improvement.

This review follows our standard approach. It has sharp, focussed findings about weaknesses in the approach taken by some statistical producers; and has clear recommendations.

But in one important respect this review differs from past reviews: it has a stronger focus on the conceptual approach we expect statistical producers to take in using administrative data; and a clearer guide to how we expect them to make the related key judgments.

We have adopted this more conceptual and guidance-focused approach because we have been surprised by the general assumption made by many statistical producers that administrative data can be relied upon with little challenge, and, unlike survey-based data, are not subject to any uncertainties. We have started from a different assumption: that statistics derived from administrative data are subject to a range of potential biases, to incompleteness and to errors.

But we also recognise that this issue has been neglected as an area for attention. As a result, it is not enough for us to simply record our concerns. We need to provide a

clear, unambiguous guide to support clear thinking and sound judgment in what for many statistical producers seems to be relatively uncharted territory.

So as a result, this review has a strong practical focus. It is built around two core insights: firstly, that not all statistics are equally risky when it comes to administrative data – many if not most statistics are low risk in terms of the quality of the underlying data. But some are higher risk, and it is important for producers to recognise this. And secondly, even for the higher risk statistics, there are a series of practices available to producers, all of which are already being deployed for some official statistics, and these practices can help provide statistical producers with necessary assurance on the data.

In this way, the Review seeks to get beyond highlighting a problem; it seeks to show that it is a problem that statisticians can address, often by drawing on existing work within their own organisations.

Finally, because of our ambition for this Review to be an authoritative statement of regulatory expectations, we want to be sure that it is complete and coherent. We are therefore publishing it initially as an Exposure Draft, on which we would welcome comments and advice. Your comments can be sent to [assessment@statistics.gsi.gov.uk](mailto:assessment@statistics.gsi.gov.uk) if possible by 30 September 2014. We hope to work with statistical producers in further developing our guidance material. Please do let us know if you are willing to be involved.

We will publish a final version once we have obtained all your comments. So we look forward to hearing from you.

**Ed Humpherson**

**Head of Assessment**

## Summary

### Introduction

1. Administrative data are a by-product of administrative systems developed primarily for operational purposes. Administrative data are used extensively in the compilation of many sets of official statistics about a wide range of topics – these include: health, such as waiting times data; crime, such as police recorded crime data; and welfare, such as the Work Programme data. As resources to fund surveys have become harder to find, technology has improved, and the demand for timely statistics has increased, the greater use of administrative data seems likely to become increasingly attractive for statistical producers.
2. However, the Authority's decision to de-designate police recorded crime as a National Statistic – because of a lack of assurance about the quality of the underlying data – has stimulated wider questioning about the accuracy and reliability of administrative data. The Authority's (March 2012) Monitoring Brief *Creating Official statistics from administrative data*<sup>1</sup> addressed the issue of statistical audit as part of a wider review.
3. This Monitoring Review considers the risks associated with the use of administrative data for statistical purposes. It identifies some examples of best practice across government in addressing those risks and presents some mechanisms for statisticians to use when seeking to implement them.

### Findings

#### 4. **Administrative data are an important source for official statistics.**

Administrative data are not collected primarily for statistical purposes. They are an increasingly common source for the production of official statistics and can be an efficient means of collating data while reducing costs to the statistical producers and to others ([paragraph 1.3 -1.4](#)).

#### 5. **There is a risk that statistical producers assume that administrative data are more reliable than survey-based data.**

An integrated theoretical framework exists for statistics based on survey data; this is not the case for administrative data, and will take some time to establish. For surveys, quality measures collated during each stage of the process are used as the basis of an explanation for users about the quality of a set of statistics. The same has to be true for statistics based on administrative data – the quality of the final product is a function of the quality checks which are

---

<sup>1</sup> <http://www.statisticsauthority.gov.uk/reports---correspondence/correspondence/letter-from-sir-michael-scholar-to-rt-hon-francis-maude---administrative-data---16032012.pdf>

carried out at all stages of the statistical process, including the design and the data compilation stage. Using administrative data is not a legitimate justification for failing to consider whether appropriate quality checks are carried out on the data and explaining to users how these affect the final statistics. Statistical producers need to build confidence in the use of administrative data for statistical purposes ([paragraph 1.5](#)).

**6. But the underlying data are subject to a range of potential biases, errors and uncertainties.**

Producers and users of survey-based statistics commonly recognise issues of uncertainty and bias in relation to survey-based statistics, and describe their scale by reporting measures such as sample size, response rates, measures of variance and precision, or descriptions of the likely sources of bias in relation to survey design and sampling. Less common, however, is the consideration of the inherent weaknesses in administrative systems, such as the effect of definitions created for local administrative purposes changing over time. We have seen evidence of statisticians having demonstrated some appreciation of the limitations of administrative data and in some cases applying good quality assurance processes to the data after they receive them. But we have also seen a lack of critical assessment of the data from administrative systems *before* they are extracted for supply into the statistical production process ([paragraph 1.6 – 1.8](#)).

**7. A range of practices are adopted by different statistical producers to provide assurance about the underlying data.**

Our review has found that statistical producers have considered these issues in a variety of ways. The case studies presented in this report demonstrate thoughtful quality assurance processes for dealing with administrative data after they are received by the statistical producer – there were good examples of process maps, sense checking and validation checks. While the case studies highlight some good practice, we have found that there is scope for further investigation into the quality of the administrative data and the circumstances in which they have been collected. Put simply, the focus of the quality assurance of administrative data needs to be widened to encompass critical thinking about the entire statistical process, including the data recording and collection stages. Just as producers monitor the entire process for survey data, similar effort should be made to understand the effect of the operational system and data processing on the quality of the administrative data. Our case studies revealed some key lessons for statistical producers: having a healthy scepticism about existing safeguards; developing constructive working relationships with data suppliers; designing management strategies for working with large numbers of data suppliers; outlining existing quality assurance processes and checks; and

seeking alternative statistical sources to provide assurance about the quality of the data ([paragraph 2.1- 2.10](#)).

**8. The degree of exposure to the risks inherent in the use of administrative data varies, and many statistics are relatively low risk.**

Not all statistics based on administrative data will require the same level of activity to provide appropriate assurance. While it is important to make visible the quality of the statistical processes and products, the degree of investigation should reflect the context in which the statistics are produced and presented. We found that it is important for statistical producers to consider the potential for data quality problems in their statistics and also the types of decisions that these statistics will inform. These considerations will allow statistical producers to determine the risks using their statistics to make those decisions; and adopt appropriate practices related to the description of this risk. We describe this conceptual framework approach as a quality assurance matrix ([paragraph 4.1 – 4.11](#)).

9. Users have told us that they understand the potential benefits of regularly collected administrative data and the potential for their contribution to official statistics. However some users have been unaware of the potential biases and uncertainty in the data which could affect how they use the statistics. Our proposed approach will provide a more secure foundation to aid users in their understanding and consideration of the judgments that they make about their use of statistics based on administrative data. We present further guidance for non-statisticians who use official statistics based on administrative data, providing some key questions that should be asked of the statistics and of those who produce them ([Annex A](#)).

## Conclusion

10. The Monitoring Review highlights the importance of statistical producers gaining and sharing with users a fuller understanding of the administrative data they use to produce official statistics, of the circumstances in which they are produced, and how they are tested and verified. In particular we also emphasise that the inherent uncertainty in the data must be communicated to the users of the statistics, to inform their use and interpretation of the statistics.

*Recommendation 1: Statistical producers should use the Quality Assurance Matrix, to determine the scale and scope of their investigations and documentation about the administrative data.*

*Recommendation 2: Statistical producers should review their quality guideline statements, Statement of Administrative Sources, and quality reports for statistics based on administrative data, to ensure that users are informed about*

*the circumstances in which administrative data are produced, of the steps taken to assure the data, and why they are satisfied that the data are sufficiently robust to be reliably used for statistical purposes.*

*Recommendation 3: In addressing recommendation 2, statistical producers should consider undertaking actions in line with the Quality Management Actions: investigate, manage and communicate model, to identify and explain to users the nature of assurance and audit arrangements associated with the administrative data and the implications for the quality of the official statistics for the most likely uses of the data.*