



**UK Statistics
Authority**

**Exposure draft of a report from the UK
Statistics Authority:**

**Quality Assurance and Audit Arrangements for
Administrative Data**

July 2014

Contents

Foreword from the Head of Assessment	4
Summary	6
Part 1: Context – the benefits and challenges of using administrative data	10
Part 2: Learning from current quality assurance and audit practice	15
Part 3: Using administrative data to produce official statistics	20
Part 4: Quality assurance and audit arrangements evaluation guide	24
Part 5: Quality assurance and audit arrangements in relation to the Code of Practice.....	32
Annexes.....	40
Annex A: Administrative data - a toolkit for non-statisticians in statistical producer bodies ..	41
Annex B: International approaches to audit arrangements and official statistics.....	44
Annex C: Case studies illustrating the use of information about the audit of administrative data by producers of official statistics.....	49

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 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*¹ 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

¹ <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.

Part 1: Context – the benefits and challenges of using administrative data

Introduction

- 1.1 This report considers the risks surrounding 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 implement when considering the quality of the data and the effect of any weaknesses on the derived statistical outputs. This report reviews: the quality checks that are carried out on administrative data before they are sent to a statistical producer; how they are questioned and examined; and how the inherent uncertainty in the data is communicated to the users of the statistics that are produced from them. The Authority recognises the resource challenges faced by statistical producers and advocates a proportional and pragmatic approach to the way that producers assess the level of assurance that is required.
- 1.2 This section presents the benefits and challenges of using administrative data in the production of official statistics. It then considers the weaknesses of administrative data and the role of quality assurance in addressing such limitations.

Use of administrative data in the compilation of official statistics

- 1.3 Administrative data are data collected for non-statistical purposes, for example, for registering births and deaths or administering benefits. It can often be personal information, for example, a person's hospital records. Administrative data can be considered as:
- (i) registration records collected for an administrative purpose, and then compiled (in principle, automatically) to form a database of administrative data (for example, birth and death records)
 - (ii) those collected for operational purposes, such as, clinical records and payments of benefits. These can be subject to differing local administrative practices and therefore might be of variable quality, especially if those tasked with collecting the data do not have a full understanding of the end purpose for the data (for example, police recorded crime statistics).
- 1.4 The use of administrative data for the compilation of official statistics has many benefits, it can: achieve cost efficiencies in terms of re-using data; allow scheduled and timely collation of data from a large number of suppliers; and reduce the response burden.

Challenges in using administrative data for statistical purposes

1.5 However, there can be limitations in the nature of administrative or operational systems that can affect the statistics derived from the underlying data. Such problems may arise from differences in definitions preferred in the statistical and operational situations, as well as changes in the operational definitions and circumstances over time. A lack of standardisation in data collection procedures, IT systems and differing local policies and priorities, can also affect the statistics. These situations require investigation by statistical producers and clear communication about the limitations to users. Both data suppliers and statistical producers need to take account of public perceptions about the use of personal data for statistical purposes² and ensure that the data are sufficiently anonymised and secure. The computational (sorting, aggregating and linking data) and inferential (identifying whether change is real, or due to chance, or to poor data quality) challenges are striking and illustrate that these contemporary concerns are evolving and dynamic. In addition, in recent years there has been considerable interest in ‘big data’³ which reflects these issues on a vastly larger scale. Box A presents a series of challenges that producers commonly face when using administrative data in the production of official statistics.

Addressing uncertainty in the data

1.6 These challenges can affect different aspects of the quality of the data, such as the reported uncertainty around the data, as well as their comparability, standardisation and coherence and enabling the linkage with other datasets. Producers and users 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. Quality measures collated during each stage of the survey process are used as the basis of an explanation for users about the quality of the statistics based on the survey data. In addition, bias may be assessed through comparison or linkage with other data sources. Less common, however, is the consideration of the inherent weaknesses in administrative or operational systems and their affect on statistics derived from them.

² Research carried out by ONS has revealed that the public expressed mixed opinions about the use of their public data for research and statistical purposes (<http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/beyond-2011-report-on-autumn-2013-consultation--and-recommendations/public-attitudes-report.pdf>). Further research carried out for the Administrative Data Research Network identified that the public are concerned when administrative data are used by other agencies http://www.esrc.ac.uk/_images/Dialogue_on_Data_report_tcm8-30270.pdf

³ ‘Big data’ typically refers to massive data sets which have the potential to reveal interesting or valuable insights into underlying processes and mechanisms which would not normally be apparent with smaller data sets. ‘Big’ can refer to the number of cases, the number of variables, the number of characteristics, the rate of data collection, or simply the complexity of the data.

Box A**Challenges using administrative data for statistical purposes:***Lack of standardised application of data collection:*

- inconsistencies in how different suppliers interpret local guidance
- differences in the use of local systems for the intended administrative function
- the distortive effects of targets and performance management regimes
- differing local priorities, data suppliers might require higher levels of accuracy for certain variables (for example payments) but less so for other aspects that are important to the statistical producer (for example demographics)

Variability in data suppliers' procedures:

- statistical producers typically do not have direct control over the development of guidance for data entry
- local checking of the data can be variable and might not identify incorrect coding or missing values
- local changes in policy could impact on how the data are recorded or on the coverage of the statistics

Quantity of data suppliers:

- there can be a large number of data suppliers, often spread geographically
- there can be many data collectors providing their data to an intermediary organisation for supply to a statistical producer

Complexity and suitability of administrative systems:

- administrative datasets can be complex containing large numbers of variables; it takes time, and therefore resource, to extract the necessary data required by the statistical producer
- data collation can be hampered by IT changes at the data supplier level
- data might need to be manipulated by the data supplier to meet the structural requirements of the statistical producer, leading to potential for errors

Public perceptions:

- lack of knowledge about use of personal data for statistical purposes
- concern that personal data should be sufficiently anonymised and secured

Quality assurance

1.7 Quality management encompasses the full range of activities carried out by statistical producers in the production of official statistics, from the initial design of data collection through to the dissemination of the statistics. A critical element of this is 'quality assurance', defined as 'the part of quality management focused on providing confidence that quality requirements will be fulfilled'⁴. Traditional quality assurance activities, such as reviewing trends or comparing data across regions, can provide statistical producers with indications of where further

⁴ International Organization for Standardization (2005): Quality management systems – Fundamentals and vocabulary (ISO 9000:2005). http://www.iso.org/iso/catalogue_detail?csnumber=42180

investigation of the underlying data could be required. Post-collection quality assurance methods, such as data validation, are an essential part of the quality assurance process, but can be of limited value if the underlying data are of poor quality. The quality of the entire statistical process directly affects the statistical products. While statisticians have demonstrated some appreciation of the limitations of administrative data, and in some cases developed good quality assurance processes after they receive the data, there has been a lack of 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 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. The Authority recognises that there are certain circumstances in which regular, systematic audit of the underlying data is essential to increase both the quality of, and public confidence in, statistics produced from administrative data.

Audit

1.8 Audit should be a key part of the administrative data quality assurance process. In this context audit means an examination of records to check their accuracy and it includes inspections and other reviews by 'neutral internal or external experts'⁵. 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, social care and statistical audit in which a sample of existing cases is investigated. These activities might be conducted on behalf of the data supplier bodies themselves as internal audit, or for regulators, such as the Care Quality Commission (CQC), or external audit or formal inspection regimes for example by the National Audit Office (NAO) or HM Inspectorate of Constabulary (HMIC). These audits should supplement, but not replace, detailed quality assurance checks carried out by statistical producers. The findings from reviews of audit arrangements will not necessarily lead to quantitative estimates of quality but can provide a richer body of evidence to inform judgments about:

- the suitability of the administrative data for use in producing official statistics
- factors the statistical producer needs to take into account in producing the official statistics
- the information that users need to know in order to make informed use of the statistics.

⁵ ESS Data Quality Management Tools paper:
http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/quality_reporting

Existing guidance

1.9 Guidance already exists for producers about the use of administrative data for statistical purposes⁶ and the National Statistician's Office has recently circulated interim guidance⁷ for producers about how to consider more carefully the quality of administrative data. In addition, there is a range of documentation available from Eurostat and some development of this topic by National Statistics Institutes (see Annex A). This review builds on this existing work.

The Authority's evaluation guide

1.10 This report highlights (in Part 2) some practices that we have identified from across the Government Statistical Service and some lessons learnt that can aid other statistical producers (fuller information is provided in Annex C). We then present evaluation guidance for statistical producers, to aid them in developing a better understanding about the quality of administrative data (in Part 3). The Authority recognises that producers are operating under tight resources; a critical aspect of addressing the concerns outlined in the paper is that statisticians take a proportionate approach based on the degree of concern about the quality of the underlying data and the public interest in the statistics – that is, the types of decisions that are informed by the statistics. The mechanisms presented in Part 4 provide statisticians with guidance on how to make these appropriate judgments. Part 5 specifies the relevant practices in the *Code* and the Authority's expectations for compliance. We conclude that section by highlighting three recommendations for statistical producers using administrative data to produce official statistics.

⁶ NSO Guidance, Use of Administrative or Management Information:

<https://gss.civilservice.gov.uk/blog/2014/05/interim-administrative-data-guidance/>

⁷ <https://gss.civilservice.gov.uk/wp-content/uploads/2012/12/Interim-Admin-Data-guidance.pdf>

Part 2: Learning from current quality assurance and audit practice

Introduction

- 2.1 In seeking to establish a standard for appropriate audit arrangements, we investigated the practices currently employed in producing six sets of official statistics. This section presents some core learning from our review of these case studies, alongside areas where we noted either existing practices, or opportunities for the development of such practices.
- 2.2 The reviews of these statistics are presented as case studies in Annex C:
- Office for National Statistics' police recorded crime statistics
 - Health and Social Care Information Centre's (HSCIC) social services activity and expenditure statistics
 - HSCIC's hospital episode statistics
 - Information Services Division's prescription statistics in Scotland
 - Department for Work and Pension's Work Programme statistics
 - Health and Safety Executive's (HSE) injury statistics
- 2.3 The case studies describe the quality assurance and audit arrangements for the selected official statistics and provide helpful examples for statistical producers to learn from others' experiences. We reviewed documentation provided for the Assessment of compliance with the *Code*, as well as published material associated with the statistics. We also drew on information from discussions with the statistical producer teams. We identified additional material about each of the administrative data sources and other relevant investigations or associated documentation such as reports published by NAO or regulator bodies.
- 2.4 We selected these case studies in order to include data from a range of different types of administrative sources and circumstances, such as multiple suppliers in local authorities and health trusts, payment systems with integrated financial audit, and an executive agency with combined responsibilities for data collection and statistics production. ONS's police recorded crime statistics are an important example as they are from an underlying administrative system with a number of established safeguards, but also one which has received considerable attention due to public concerns about the accuracy of the data and the limited assurance provided by the safeguards.
- 2.5 We now describe some lessons revealed by the case studies that illustrate some ways that the practices have been used to better understand the quality of statistics derived from administrative data.

Lesson 1: Don't trust the safeguards

2.6 It is clear from the police recorded crime case study (C1) that a formalised process for the management of administrative data and its assurance may not be enough in itself. The environment in which the organisations are operating may also play a role in influencing the handling of information and the wider context needs to be understood by statistical producers. For example, the producers should identify the potential distortive effect of targets and performance management regimes on the behaviour of those compiling the administrative information. These pressures should be addressed by the data supplier bodies with further appropriate safeguards taken. Statistical producers need to identify the associated weaknesses of the data and determine whether the safeguards are functioning effectively, and what more might be done to raise the level of assurance. They should ensure that they communicate clearly the implications for the statistics in relation to their use.

Lesson 2: Get involved

2.7 HSCIC has established strong ties with local authorities (LAs) in England with responsibility for adult social care services (case study C2). It has arranged secondments of LA staff to work in HSCIC. This opportunity ensured that a good understanding of the operational context underpinned the development of the new data collections by HSCIC, after a thorough review of adult social care data and users' needs. The importance of establishing clear agreements and relationships between statistical producers and data suppliers is also highlighted in the National Statistician's guidance on the use of administrative data.

Lesson 3: Raise a red flag

2.8 HSCIC's Hospital Episode Statistics case study (C3) illustrates the challenge of having an extremely large number of data collectors and suppliers – there are around 700 health trusts. In order to make the investigation of data quality concerns manageable, HSCIC draws on its effective engagement with supplier representative bodies and information governance groups, to identify potential data quality concerns. It also uses its own quality assurance of the supplied data to identify suspected issues or 'red flags'. For example, HSCIC reviews data for all the health trusts to identify whether any data suppliers are outliers that require further investigation. Audit information about the practices of the individual trusts may provide additional evidence for statistical producers when seeking to determine the adequacy of operational arrangements. The compilation of performance indicators across the supplier organisations, for example, for benchmarking, could also be used by statistical producers to further corroborate the effectiveness of the organisational arrangements in the supplier bodies. These types of evidence present an opportunity for further investigation to

support the statistical team's judgment about the suitability of the organisations' practices in producing and supplying the administrative data. HSCIC recognises the importance of establishing the quality of the data to meet operational or clinical needs, as well as for the secondary use of producing official statistics. Sharing this type of information across relevant statistical teams would help the statistical producer gain maximise benefit from the investigations.

Lesson 4: See the big picture

2.9 Central to judging the appropriateness of audit arrangements is to identify what specific audits and investigations have been conducted, how frequently and by whom. The case studies on ISD's prescription statistics (C4) and DWP's Work Programme statistics (C5) are based on payment information, and financial audit is an inherent part of the checks conducted on the data. Both case studies also include examples of process maps, with the ISD example indicating the points at which checks are made. The DWP case study demonstrates the detailed checks made to ascertain the suitability of the administrative source for use in producing official statistics. It also places the financial audit within the context of the quality assurance checks carried out by the statistical team.

Lesson 5: Corroborate the evidence

2.10 HSE (C6) has identified an under-recording of injuries in its non-fatal injury data, and so commissions questions on the Labour Force Survey to measure self-reported injury. This example highlights the fact that weaknesses in some administrative data sources cannot necessarily be overcome. Statisticians may need to seek alternative information to corroborate the administrative data. HSE uses different types of checks to quality assure the injury data: thorough checks by the HSE Inspectors, built-in system checks when employers notify HSE of an injury, internal audit of the system itself and its subsequent improvement, and the use of sample audit to check the completeness of the injury data. The audit of sampled data was found to be a particularly important source of evidence, to support the use of the administrative data, both for DWP and ISD, as well as HSE.

Conclusion

2.11 We identified some areas of good practice in the assurance and auditing of administrative data, as well as areas where there are opportunities for statistical producers to provide further information about quality assurance and audit arrangements – see Table 2.1 for a summary. We found that there is often a great deal of existing information available that is not accessed or used by the statistical producers to assess the quality of the data. A number of the practices identified in the case studies are specifically mentioned in the National

Statistician's existing guidance⁸ about the use of administrative data – this guide emphasises the need to investigate the data source, work closely with data suppliers, and understand possible causes of error through validation and triangulation with other sources. Determining the suitability of the data should not be considered a one-off judgment but an ongoing process of challenge. Concerns identified through statistical producers' own quality assurance checks provide an opportunity to identify weaknesses or risks within the data supply process. New issues can emerge as operational changes occur to the administrative systems, as well as the implementation of classification and coding changes.

- 2.12 Each of the producers has developed detailed checks on the data received from the suppliers. Having identified anomalies, they request that the data suppliers investigate further the issues and possibly resubmit the data. Very often, though, the investigations focus solely on the internal validity of the data. We found that statistical producers tend not to use the quality issues they have found as prompts to probe the data suppliers' audit arrangements, or identify systemic or organisational issues that contribute to poor data.
- 2.13 In Part 3 we extend the areas of good practice identified in the six case studies to suggest a conceptual framework for evaluating the assurance and audit arrangements of administrative data in ways that are consistent with the *Code*. We then propose some tools to help statistical producers make this critical evaluation of the administrative data. These require the producer to first gauge the level of concern about data quality and the importance of the statistics derived from the administrative data. A maturity matrix (the '**Quality Assurance (QA) Matrix**') then guides the producer in determining the scale of investigation and documentation that are consistent with the data quality concerns and the nature of the use of the statistics.

⁸ <https://gss.civilservice.gov.uk/wp-content/uploads/2012/12/Interim-Admin-Data-guidance.pdf>

Table 2.1: Examples of audit and assurance practices arising from the six case studies

Practice	Case study practice example
Determine the capacity and capability of the data supplier organisation(s) to provide accurate and reliable data:	
Gather information about assurance and audit arrangements from external data suppliers	ISD's prescriptions statistics
Develop a clear process map	ISD's prescriptions statistics
Identify audit and assurance issues at institutional and operational levels	HSE's injury statistics
Identify the potential impact of targets and performance management regimes	ONS's police recorded crime statistics
Identify the results of internal audit of the systems and processes for administrative data	DWP's Work Programme statistics
Identify the frequency and nature of external oversight for high profile sets of statistics	ONS's police recorded crime statistics
Collate organisational/operational information in a form that is readily accessible by statisticians in making judgments about the suitability and accuracy of the administrative data	HSCIC's HES statistics
Develop co-operative arrangements with data supplier organisation(s)	HSCIC's social services statistics
Identify how clearly the roles of those who record, quality assure and sign off the data for delivery are specified and understood	ONS's police recorded crime statistics
Establish whether a Memorandum of Understanding or a Service Level Agreement exists between the statistical producers and data suppliers:	
Identify how clearly the roles and responsibilities of those involved in the process are specified	ONS's police recorded crime statistics
Determine whether the key points from this agreement have been published	ONS's police recorded crime statistics
Fully document audit arrangements and identify the implications for the statistics:	
Explain each stage of data collection, processing and quality assurance, demonstrating who is responsible at each stage and the checks that they carry out on the data	HSCIC's HES statistics
Identify the potential risks to the accuracy of the data and the safeguards that are in place to minimise the risk	DWP's Work Programme statistics
Identify alternative data and information sources e.g. from surveys – to verify accuracy of the data	HSE's injury statistics

Part 3: Using administrative data to produce official statistics

Introduction

- 3.1 The previous part of this report outlined the main lessons from a review of current approaches to the quality assurance of official statistics based on administrative data. These lessons informed our development of a conceptual framework that describes the range of practices that producers should consider when using administrative data for official statistics. Part 3 explains this practice model.
- 3.2 The *Code* requires statistical producers to ensure that administrative sources are fully exploited for statistical purposes, with appropriate safeguards in place. Statistical producers must ensure that they use data that are based on definitions and concepts which approximate well those required for the statistics, and that the quality of the data is sufficiently robust. The *Code* also requires producers to inform users about the quality of their statistical outputs, including estimates of the main sources of bias and other errors in the data. This supporting metadata should include information about the quality assurance procedures and the arrangements for auditing the quality of the data. However, the steps to be taken by statistical producers need to go beyond a narrow interpretation of ‘quality assurance’; they also encompass the working arrangements and relationships with the other agents, particularly data suppliers.
- 3.3 The practice model that we propose sets out four areas of practice in relation to the *Code* (see Figure 3.1 below):
- Operational context and administrative data collection
 - Communication with data suppliers
 - Suppliers’ quality assurance principles, standards and quality checks
 - The producer’s quality assurance investigations and documentation

Operational context and administrative data collection

- 3.4 *Operational context* reflects the need for statistical producers to gain an understanding of the environment and processes in which the administrative data are being compiled and the factors which might increase the risks to the quality of the administrative 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 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 the safeguards.

3.5 The *administrative 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 administrative systems as being 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.

Communication with data suppliers

3.6 *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 administrative systems and documented data supply arrangements. When multiple data suppliers are 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. Producers should also determine whether specific data quality indicators are relevant and can be provided by data suppliers. ESS quality guidelines⁹ highlight a number of quality indicators relevant to administrative data:

- Data completeness – are required data variables supplied?
- Over-coverage – are units outside the target population included?
- Unit non-response – are there whole units with no (usable) information?
- Item non-response – are particular variables missing information?

Suppliers’ Quality Assurance principles, standards and quality checks

3.7 Statistical producers should understand the validation checks that are conducted by the supplier, and the results of the checks. Some operational systems will

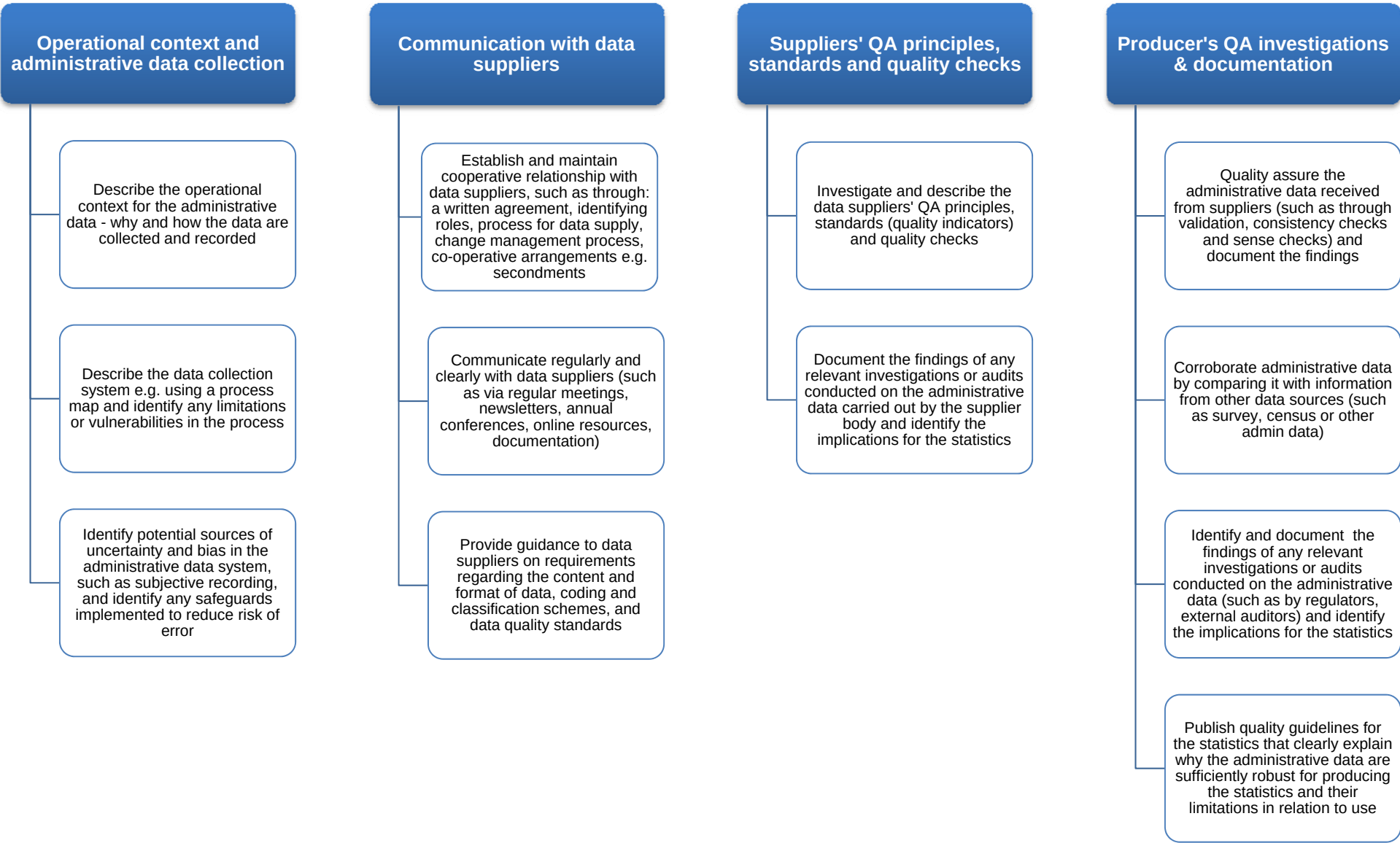
⁹ ESS Guidelines for the implementation of the ESS Quality and Performance Indicators 2014:
http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/quality_reporting

also have a process of audit established – 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 administrative data, that is, the closeness of computations or estimates to the true values, as well as its validity.

The producer's Quality Assurance investigations and documentation

- 3.8 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. The checks conducted on data received from data suppliers are well established and represent the main body of work undertaken by producers to verify the validity of the data prior to use in producing official statistics. Since the checks cannot, by themselves, verify the accuracy of the administrative data, producers should seek additional information. They should corroborate their quality assurance findings against data from other sources, such as surveys or other administrative 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, auditors, or by professional bodies (such as Local Government Association).
- 3.9 The findings from the producer's quality assurance checks should be supplemented by the knowledge gained through reviewing the other practice areas outlined above, to inform a published statement that sets out the basis of the producer's judgment about the quality of the administrative data.

Figure 3.1: Practices to be undertaken by statistical producers when using administrative data



Part 4: Quality assurance and audit arrangements evaluation guide

4.1 In Part 3 we set out our practice model that describes the four areas of practice for producers using administrative data for statistical purposes. In Part 4 we present a maturity matrix, grounded in the practice model, to help statistical producers determine the scale of investigation and documentation required to be assured that the quality of the data is suitable, and to provide informed assurance to users.

4.2 The approach that we propose here has two parts:

- a) *Data Quality Concern and Public Interest Profile Matrix* – producers can use this matrix to decide the characteristics of their statistics, and make decisions that are pragmatic and proportionate.
- b) *Quality Assurance Matrix* – producers can use this maturity matrix to identify the appropriate level of assurance to be undertaken and documentation to be published.

Data quality concern and public interest profile matrix

4.3 Judgments about the quality of the data for use in official statistics should 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. We have summarised the relationship between these two dimensions in the quality concern and profile matrix (Table 4.1) below. It may be the case that the circumstances surrounding the statistics change which, in turn, may require this evaluation to be repeated.

Table 4.1: Data quality concern and public interest profile matrix

Level of concern over data quality	Public interest profile: importance for informing decisions		
	Lower	Medium	Higher
Lower level ('low')	Statistics of lower data quality concern and lower public interest [M1]	Statistics of lower data quality concern and moderate public interest [M1/M2]	Statistics of lower data quality concern and high public interest [M1/M2]
Medium ('medium')	Statistics of moderate data quality concern and lower public interest [M2]	Statistics of moderate data quality concern and moderate public interest [M2]	Statistics of moderate data quality concern and high public interest [M2/M3]
Higher level ('High')	Statistics of high data quality concern and lower public interest [M3]	Statistics of high data quality concern and moderate public interest [M3]	Statistics of high data quality concern and high public interest [M3]

M1 to M3 = the three maturity levels in the QA Matrix

- 4.4 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. Table 4.2 below provides further examples of the criteria for judging the degree of concern with data quality.
- 4.5 The public interest profile reflects the importance of the decisions informed by the statistics. Higher public interest 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.
- 4.6 The two dimensions are described below in Tables 4.2 and 4.3, with reference to example case studies for illustrative purposes.

Table 4.2: Level of concern over the accuracy of the data

Examples of the criteria for considering degree of concern over data quality

Level of concern about data quality	Criterion
Lower	Single data supplier
	Simple data collection process
	Well-defined classifications
	Clear coding frame
	Clear instructions for recording
	Validation checks built into data collection system
	Validation checks built into statistical producer's system
	Internal or financial audit part of operational checks
	Well defined roles and responsibilities
	No performance management regime or use of targets
	International standards for measurement
	External oversight/audit (e.g. by regulators such as NAO, Ofqual, CQC)
	<i>Case study example: ISD's prescription statistics - single provider (Prescribing Services Division) who has built-in validation system checks and financial audit into the operational process, with clearly defined roles and data management arrangements</i>
Medium	Combination of factors from lower and higher levels with safeguards to moderate the concerns
	<i>Case study example: DWP's Work Programme statistics - multiple providers, payment by results offset by integrated financial audit and operational checks</i>
Higher	Multiple data supply and/or data collection bodies
	Complex data collection
	Subjective recording of variables
	Lack of consistency in coding
	Lack of clarity in classification systems
	No audit of administrative data within operational processes
	Over-reliance on system checks for checking accuracy of data
	Poorly defined roles and responsibilities
	Performance management regime
	Use of targets – possibility of distortive effects on the statistics
	Lack of external oversight
	<i>Case study example: ONS's police recorded crime statistics - multiple data suppliers with variable recording practices, subjective interpretation of information, use of local targets to drive performance, lack of external scrutiny, indirect relationship between statistical producer and data suppliers with intermediary statistical producer (Home Office) receiving and processing administrative data</i>

Table 4.3: Public interest profile of the statistics

Examples of the criteria for considering the degree of public interest and use in informing decision making

Profile level	Criterion
Lower	Always likely to be a politically neutral subject
	Interest limited to niche user base
	Not economically sensitive
	Limited media interest
Medium	Wider user interest
	Moderate economic and/or political sensitivity
	<i>Case study example: HSCIC's adult social services activity statistics - public interest in adequacies of social care services</i>
Higher	Legal requirement, for example, for Eurostat
	Economically important, reflected in market sensitivity
	Substantial level of resource and key to that allocation
	High political sensitivity, reflected by Select Committee hearings
	Substantial media coverage of policies and statistics
	Public health issue
	<i>Case study example: ONS's police recorded crime statistics - strong public interest in crime, fear of crime, effectiveness of police and holding Government to account, strong public concern in mis-recording of crime by police forces, investigation by Public Administration Select Committee</i>

Quality Assurance Matrix

- 4.7 Having assessed the levels of data quality concern and public interest, statistical producers can use the maturity matrix, called the **Quality Assurance (QA) Matrix**, to determine the appropriate scale of assurance and documentation required to inform themselves and users about the quality assurance and audit arrangements for the administrative data. This will support the basis of their judgment of the suitability for the use of the data for statistical purposes.
- 4.8 The **maturity levels diagram** (Figure 4.2 below) explains the different levels of assurance that are used in the QA Matrix – the requirement for investigation and documentation increases at each level from ‘basic’ (M1) to ‘enhanced’ (M2) to ‘comprehensive’ (M3).
- 4.9 The QA Matrix (Table 4.4 below), outlines the types of practices that can be undertaken to provide assurance of each aspect of the quality assurance and audit arrangements and to document the judgment. Producers should decide which of the maturity levels is appropriate for the administrative data under consideration for the four practices. Producers must explain the basis of their judgments of the level of assurance.
- 4.10 The Authority does not regard ‘No assurance’ (M0) as compliant with the *Code*.
- 4.11 During the Authority’s Assessment of official or National Statistics based on administrative data (of compliance with the *Code*), the assessors will also make an evaluation of what they regard as the appropriate maturity levels for the administrative data. The Authority may consider that, given the level of concern over data quality and the public interest profile of the statistics, a higher level of maturity (M2 or M3) is appropriate than that judged by the statistical producer. In these cases, applying the measure associate with either M1 or M2 may be viewed as not compliant with the *Code*.

Figure 4.2: Maturity levels: degree of scrutiny and documentation to be carried out by statistical producers

M1: Basic Assurance

- *Statistical producer reviews the administrative data QA arrangements and publishes a high-level summary of the assurance*
- In general, statistics of **lower data quality concerns** and **lower public interest** will adhere to this maturity level
- This level may also be appropriate for statistics of:
 - **lower** data quality concerns and **moderate** public interest
- In some rarer circumstances, this level may be appropriate for statistics with:
 - **lower** data quality concerns and **high** public interest

M2: Enhanced Assurance

- *Statistical producer evaluates the administrative data QA arrangements and publishes a fuller description of the assurance*
- In general, statistics of **moderate data quality** and **moderate public interest concerns** will adhere to this maturity level
- This level may also be appropriate for statistics of:
 - **lower** data quality concerns and **moderate** public interest
 - **moderate** data quality concerns and **lower** public interest
 - **moderate** data quality concerns and **high** public interest

M3: Comprehensive Assurance & Audit

- *Statistical producer investigates the administrative data QA arrangements and the results of independent audit and publishes detailed documentation about the assurance*
- In general, statistics with **high data quality concerns** will adhere to this maturity level:
 - **high** data quality concerns and **lower** public interest
 - **high** data quality concerns and **moderate** public interest
 - **high** data quality concerns and **high** public interest
- It could also be relevant where the data quality concerns are less but there is a **high public interest** in the statistics.

Table 4.4: Quality Assurance Matrix

Maturity level	Areas of practice related to quality assurance and audit	
	A: Operational context and administrative data collection	B: Communication with data suppliers
M0: No assurance <i>Not compliant with the Code of Practice</i>	No documentation of operational context and administrative data collection by supplier	No documentation of data supply agreement, roles and responsibilities
M1: Basic assurance <i>Statistical producer reviews the administrative data QA arrangements and publishes a high-level summary of the assurance</i>	- Outline administrative data collection process - Process map of the administrative data collection - Outline operational context - Identify and summarise the implications for accuracy and quality of data - Identify safeguards taken to minimise risks to data quality	- Basic communication, such as: - annual statement of needs - timing and format of data supply - coordination of data sign off from data suppliers - agreed feedback process of identified errors to data suppliers and recording of data supplier response
M2: Enhanced assurance <i>Statistical producer evaluates the administrative data QA arrangements and publishes a fuller description of the assurance</i>	- Fuller description of operational context and administrative data collection, such as: - more detailed process map explaining data collection processes, - explanations for classifications, - Identify and summarise potential sources of bias and error in administrative system - Identify and describe safeguards taken to minimise risks to data quality - More detailed description of the implications for accuracy and quality of data	- Clear mode of communication - Specify timing, form and content for data supply - Security and confidentiality protection - Regular engagement with suppliers: - assign Single Point of Contact role for both producers and data suppliers
M3: Comprehensive assurance & audit <i>Statistical producer investigates the administrative data QA arrangements and the results of independent audit, and publishes detailed documentation about the assurance and audit</i>	- Detailed description of administrative system and operational context: - explain why the data are collected, who by and how - identify differences across areas in collection and recording of the data, - identify issues for individual data items - whether objective or based on subjective recording, - identify issues in design, definition of targets - Detailed and specific description of the implications for accuracy and quality of the data, - Identify and explain any safeguards used to minimise the risks to data quality	- Establish/maintain cooperative relationship, - Written agreement specifying: - roles and responsibilities, - data supply process, - schedule, - content specification; - Establish change management process; - Communicate regularly, e.g. through meetings, newsletters, conferences - Attend data supplier group meetings - Secondments

Table 4.4 (continued): Quality Assurance Matrix

Areas of practice related to quality assurance and audit		Maturity level
<i>C: Suppliers' QA principles, standards and checks</i>	<i>D: Producer's QA investigations and documentation</i>	
• No description of suppliers' QA procedures and standards	• No description of own QA checks	M0: No assurance <i>Not compliant with the Code of Practice</i>
• Some knowledge of suppliers' QA checks with brief description, • Identify whether audits are conducted on the admin data, • Describe the implications for the statistics	• Some description of own QA checks on the admin data, • Outline general approach and overall findings, • Identify the strengths and limitations of the admin data, • Explain the likely degree of risk to the quality of the admin data provided by the operational context and data collection approach	M1: Basic assurance <i>Statistical producer reviews the administrative data QA arrangements and publishes a high-level summary of the assurance</i>
• Description of suppliers' QA principles, standards (quality indicators), and checks; • Identify and describe the audit of admin data, • Describe the implications for the statistics	• Provide a fuller description of own QA checks on the admin data, • Detail the general approach and findings for specific quality indicators; • Identify the strengths and limitations of the admin data; • Explain the likely degree of risk to the quality of the admin data provided by the operational context and data collection approach	M2: Enhanced assurance <i>Statistical producer evaluates the administrative data QA arrangements and publishes a fuller description of the assurance</i>
• Describe the data suppliers' principles, standards (quality indicators) and quality checks; • Identify and review quality reports for the data; • Identify and document the findings of investigations and audits conducted on the admin data and associated targets • Describe the implications for the statistics	• Provide a detailed description of own QA checks on the admin data, • Give quantitative (and where appropriate qualitative) findings for specific quality indicators; • Undertake comparisons with other relevant data sources (such as survey or other admin data); • Identify possible distortive effects of targets • Identify the strengths and limitations of the admin data and any constraints on use for producing statistics; • Explain the likely degree of risk to the quality of the admin data provided by the operational context and data collection approach	M3: Comprehensive assurance & audit <i>Statistical producer investigates the administrative data QA arrangements and the results of independent audit, and publishes detailed documentation about the assurance and audit</i>

Part 5: Quality assurance and audit arrangements in relation to the *Code of Practice*

Introduction

- 5.1 Part 4 described how the QA Matrix can guide statistical producers in judging the amount of investigation and documentation that is necessary to adequately describe the assurance and audit arrangements of the administrative data. Part 5 relates the practices identified for the appropriate quality assurance of the data to the *Code of Practice for Official Statistics*. It also emphasises the importance of ensuring that the activities to assure the data are part of the producer's broader quality management practice and thinking.
- 5.2 The *Code* contains a variety of practices that reflect the need to determine and explain the nature and level of the quality of administrative data. Some are organisational aspects of the statistical producer body – for example, explaining its approach to quality management, and publishing quality guidelines. Others reflect the quality assurance arrangements adopted by statistical teams, and the importance of close working relationships with data suppliers. The practices are listed in Table 5.1 (at the end of the section).

Applying the *Code*

- 5.3 As illustrated in the practice model in Part 3, developing an understanding of audit arrangements for administrative data should be considered as part of a producer's overall approach to quality assurance. However, explaining the nature of these arrangements and their implications for the statistics does not necessarily require producers to release separate material from the supporting information currently published. Instead it would be most helpful to users for the information to be provided within existing quality material, with clear signposting between relevant documents.

(a) Quality Guidelines¹⁰

- 5.4 The *European Statistical System Quality Assurance Framework (ESS QAF)* highlights the need for published quality guidelines that set out how a statistical producer implements quality management. The guidelines should include a description of the statistical production processes and the methods used to monitor the quality at each stage of the process. Similarly the practice model in Part 3 of this report highlights the importance of providing insight into the data collection and operational circumstances in which administrative data are

¹⁰ Principle 4, Practice 4 of the *Code of Practice*

produced, as they may affect the quality of the underlying data and the statistics derived from them.

(b) Statement of Administrative Sources¹¹

- 5.5 The Statement of Administrative Sources (SoAS) requires producers to explain the audit arrangements for administrative data used for statistical purposes. Producers often provide a general statement reflecting the organisational approach to administrative data, with specific information provided for individual sets of statistics in their accompanying supporting material. As long as the signposting to the assurance and audit information is provided and is clear, this presentation of the SoAS is compliant with the *Code*.
- 5.6 Reviewing the assurance and audit arrangements for the administrative data and extending quality information will present a useful opportunity for producers to review the information provided in their SoAS. The SoAS can be used to explain the approach taken by groups of data collectors and suppliers (such as by NHS and Foundation trusts, local authorities and police forces) to assure themselves that the administrative data are both validated and their accuracy verified. These approaches may reflect the internal and external audit by the organisations themselves and their regulators. It may also describe (or signpost users to) the benchmarking comparisons by professional bodies, such as given by Local Government Association for local authorities on its Inform website¹².

(c) Quality Reports¹³

- 5.7 Statistical producers are also required to publish information about the quality and reliability of statistics to accompany the official statistics. The *ESS Handbook for Quality Reports* provides guidance to producers about the content of the guidelines. It recommends that these be based on the statistical output quality dimensions of the *ESS Quality Assurance Framework*: relevance, accuracy, comparability, coherence, timeliness and accessibility.
- 5.8 The ESS Handbook highlights that producers should ensure that the section on 'relevance' presents information about the organisational and operational circumstances surrounding the collection of the administrative data. These descriptions would be particularly helpful for users if they were based on a process map which illustrates the various elements of the data collection and supply process, and they should highlight the safeguards that are in place to minimise errors in data recording and processing. The section on 'accuracy' should provide information about the quality assurance and audit of the administrative data drawn from the investigations conducted by both the data

¹¹ Protocol 3, Practice 5 of the *Code of Practice*

¹² <http://www.local.gov.uk/about-lginform>

¹³ Principle 4, Practice 2 and Principle 8, Practice 1 of the *Code of Practice*

supplier(s) and the statistical producer. It would be good practice to provide specific quality indicators, such as rates of missing data and under-coverage. The process map could also be used to highlight the quality indicators.

Quality Management Actions

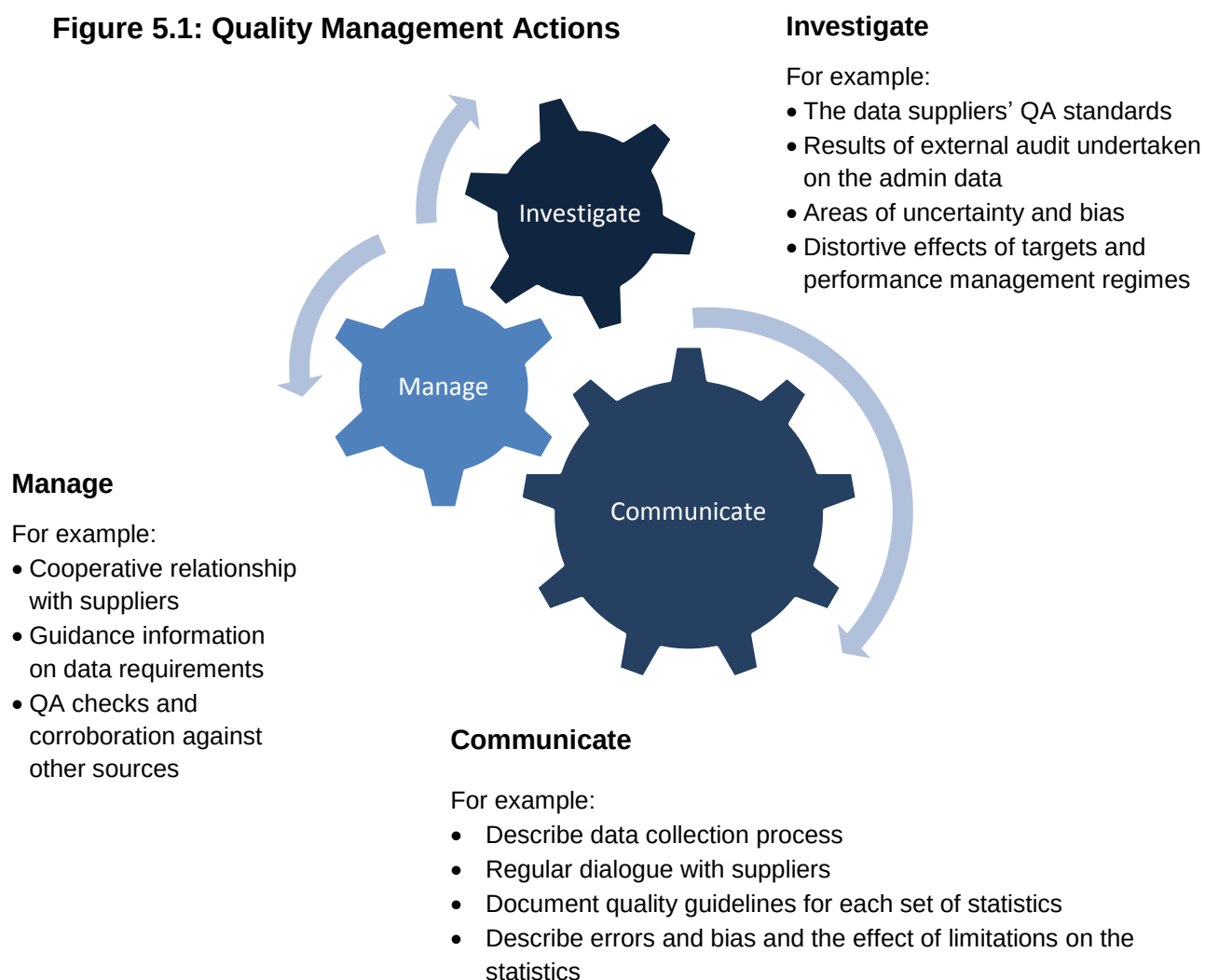
5.9 Statistical producers can use the maturity matrix (Part 4) to determine the level of detail required for specific sets of statistics, given their level of importance in informing decision making and the degree of concern about the accuracy of the underlying data. Having identified which maturity level is appropriate for each of the four practice areas, we recommend that producers carry out the following quality management actions:

- Investigate
- Manage
- Communicate

5.10 These represent three types of actions for assuring the quality of the administrative data and in documenting the findings. They draw on the practices highlighted in the four practice areas of the practice model. Producers should **investigate** – for example, the types of checks carried out by data collectors and data suppliers, as well as the operational circumstances in which the data are produced. They should also identify potential sources of bias in the production process. Producers should **manage** their relationships with the data suppliers – by establishing clear processes for data supply and for managing change. They should also maintain regular quality assurance checks of the data and use other data sources where possible to corroborate their findings. And producers should **communicate** clearly with their data suppliers and their users – to ensure a good understanding of the strengths and limitations of the administrative data.

5.11 As reflected by the cog diagram below (Figure 5.1), these are practices that are continuous and iterative, reflecting the ongoing use of the data and the dynamic nature of operational environments. Producers should not regard their review as a one-off event, but rather as a process that requires repeated evaluation, to understand the implications of changes and allow for the ongoing monitoring of the quality of the administrative data.

Figure 5.1: Quality Management Actions



Conclusion

5.12 This report highlights the importance of statistical producers gaining a fuller understanding of the administrative data that they use for deriving official statistics, of the circumstances in which the data are produced, and how they are tested and verified. We also emphasise that the inherent uncertainty in the data is communicated to the users of the statistics, to aid their use and interpretation of the statistics.

5.13 We recommend that statistical producers:

Recommendation 1:

Use the Quality Assurance Matrix, to determine the scale and scope of their investigations and documentation about the administrative data.

Recommendation 2:

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 used for statistical purposes.

Recommendation 3:

In addressing recommendation 2, consider undertaking actions in line with the Quality Management Actions: investigate, manage and communicate, 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.

Table 5.1: Excerpts from the *Code* relevant to the assurance and audit of administrative data for each practice area

Principle or Protocol and specific practice number	Relevant parts of the practice	Areas of practice
Principle 4.1	Publish details of the methods adopted, including explanations of why particular choices were made	<i>1: Operational context & administrative data collection</i>
Principle 4.2	Ensure ... that users are informed about the quality of statistical outputs, including estimates of the main sources of bias and other errors	<i>1: Operational context & administrative data collection</i>
Principle 4.3	Adopt quality assurance procedures	<i>3: Suppliers' QA principles, standards and quality checks</i>
		<i>4: Producer's QA investigations and documentation</i>
Principle 4.4	Publish quality guidelines	<i>3: Suppliers' QA principles, standards and quality checks</i>
		<i>4: Producer's QA investigations and documentation</i>
Principle 4.5	Seek to achieve continuous improvement in statistical processes	<i>2: Communication with data suppliers</i>
		<i>3: Suppliers' QA principles, standards and quality checks</i>
		<i>4: Producer's QA investigations and documentation</i>
Principle 4.6	Promote comparability... by adopting common standards, concepts ... definitions, statistical units and classifications...	<i>2: Communication with data suppliers</i>
		<i>4: Producer's QA investigations and documentation</i>
Principle 4.7	Where time series are revised, or changes are made to methods or coverage, produce consistent historical data where possible	<i>2: Communication with data suppliers</i>
		<i>4: Producer's QA investigations and documentation</i>
Principle 6.3	Promote statistical purposes actively in the design of administrative systems in order to enhance the statistical potential of administrative records	<i>2: Communication with data suppliers</i>
Principle 6.4	Analyse the costs of proposed new data requirements (to data suppliers) against the potential benefits	<i>2: Communication with data suppliers</i>
Principle 6.5	Evaluate existing data sources and estimation techniques before undertaking new surveys	<i>2: Communication with data suppliers</i>
Principle 7.1	Ensure that statistical services have the	<i>2: Communication with data</i>

	staff, financial and computing resources to produce, manage and disseminate official statistics to the standards of this Code	<i>suppliers</i>
Principle 7.4	Monitor expenditure against work programmes and demonstrate effective stewardship of resources allocated to statistical work	<i>2: Communication with data suppliers</i>
Principle 7.5	Seek to balance quality (for example, accuracy and timeliness) against costs (including both costs to government and data suppliers), taking into account the expected uses of the statistics	<i>2: Communication with data suppliers</i>
		<i>3: Suppliers' QA principles, standards and quality checks</i>
		<i>4: Producer's QA investigations and documentation</i>
Principle 7.6	Ensure that appropriately skilled people are employed in the statistical production process.	<i>4: Producer's QA investigations and documentation</i>
Principle 7.7	Where administrative data are used for statistical purposes, follow the practices set out in Protocol 3	<i>4: Producer's QA investigations and documentation</i>
Principle 8.1	Provide information on the quality and reliability of statistics in relation to the range of potential uses, and on methods, procedures, and classifications	<i>1: Operational context & administrative data collection</i>
		<i>3: Suppliers' QA principles, standards and quality checks</i>
		<i>4: Producer's QA investigations and documentation</i>
Principle 8.2	Provide factual information about the policy or operational context of the official statistics	<i>4: Producer's QA investigations and documentation</i>
Protocol 1.4	Provide users with information about the quality of the statistics, including any statistical biases	<i>1: Operational context & administrative data collection</i>
		<i>3: Suppliers' QA principles, standards and quality checks</i>
		<i>4: Producer's QA investigations and documentation</i>
Protocol 3.2	Only base statistics on administrative data where the definitions and concepts are good approximations to those appropriate for statistical purposes	<i>2: Communication with data suppliers</i>
		<i>3: Suppliers' QA principles, standards and quality checks</i>
Protocol 3.3	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 information. Where possible, use common information technology and information management systems that facilitate the flow of information between producers of statistics	<i>2: Communication with data suppliers</i>
		<i>3: Suppliers' QA principles, standards and quality checks</i>
		<i>4: Producer's QA investigations and documentation</i>

Protocol 3.4	Ensure that no action is taken within the producer body, or public statement made, that might undermine confidence in the independence of the statistics when released	<i>2: Communication with data suppliers</i>
		<i>3: Suppliers' QA principles, standards and quality checks</i>
Protocol 3.5	Prepare ...a Statement of administrative sources which identifies the following:	<i>3: Suppliers' QA principles, standards and quality checks</i> <i>4: Producer's QA investigations and documentation</i>
	a) The administrative systems currently used in the production of official statistics	
	b) Procedures to be followed within the organisation to ensure that full account is taken of the implications for official statistics when changes to administrative systems are contemplated	
	c) Information on other administrative sources that are not currently used in the production of official statistics but have potential to be so used	
	d) Arrangements for providing statistical staff, whether inside the producer body or elsewhere, with access to administrative data for statistical purposes	
	e) Arrangements for auditing the quality of administrative data used for statistical purposes	
	f) Arrangements for ensuring the security of statistical processes that draw on administrative data	

Annexes

Annex A

Administrative data - a toolkit for non-statisticians in statistical producer bodies

Annex B

International Approaches to Audit Arrangements and Official Statistics

Annex C

Case Studies Illustrating the Use of Information about the Audit of Administrative Data by Producers of Official Statistics:

- C1 ONS's Police Recorded Crime Statistics
- C2 HSCIC's Social Services Activity Statistics
- C3 HSCIC's Hospital Episode Statistics
- C4 ISD's Prescription Statistics in Scotland
- C5 DWP's Work Programme Statistics
- C6 Health and Safety Executive's Injury Statistics

Annex A: Administrative data - a toolkit for non-statisticians in statistical producer bodies

The purpose of this toolkit is to provide a framework to guide the judgments of non-statisticians when confronted with statistics derived from administrative data.

Administrative data are the lifeblood of any organisation. They record what has been done, by whom, to whom, when and where. They range from customer records to payment details, from records of activity like the number of crimes recorded by police to records of geography and the environment like the mapping of flooding incidence by environmental regulators.

Administrative data are frequently, and increasingly, used as the basis for official statistics. This toolkit seeks to support policy makers when they encounter administrative data. Policy makers in government bodies will come across administrative data in different ways. These include when their organisation is responsible for the production of official statistics, and our questions have this primary use in mind. But policy makers may also be presented with administrative data when confronting a new policy problem, and they want to see what their organisation already knows about the problem; and when they are evaluating the success of a policy as it is implemented. The questions below will be a useful guide in those situations too.

While most of these questions are relevant to the other main source of official statistics – surveys – they are particularly pertinent to statistics based on administrative data because they are a relatively less well recognised feature of the statistical system. We are therefore seeking to help policy makers become more intelligent customers through this toolkit.

The toolkit takes the form of 10 key questions. While they work as a sequence of questions, there is no need to work through in methodical order; any of these questions can be asked on its own or in combination with any of the others. The key point is for the policy maker to be willing to interrogate the strength of the data and obtain assurance on the reliability of that data.

The 10 questions

1. Where do the data come from?

This is a simple, straightforward question. Your statisticians should be able to give you a clear answer, explaining the nature of the data, who produces them and why. This question is a good starting point – and if the answer sounds vague or unconvincing, this can be the jumping off point for further questions as you look to interrogate the data more rigorously.

2. Is there a consistent time series?

A good way of understanding the statistics is to see how they behave over time. A time series can demonstrate to you how the underlying data have been affected by contextual factors – which you may know better as a policymaker than the statisticians who work on the production of the official statistics.

3. If there's a limited time series, how do you caveat the statistics you've got, and what warnings do you give about the conclusions that can be drawn?

This is crucial. Your organisation is likely to be criticised – not least by the UK Statistics Authority – if you disseminate statistics which argue for a clear pattern or response to a policy initiative when you don't have a time series to enable you to make these claims with confidence. It's always worth considering the extent to which the time series might be subject to cyclical patterns, such as economic growth.

4. What is the story behind the pattern revealed by the statistics?

Understanding the story behind the patterns is important – though this is more an internal check: to satisfy yourself that the patterns are plausible and fit your own experience. Be cautious about going public with this story until you have explored other explanations for the patterns in the data.

5. Have you changed measurement – or data suppliers – and might this be a plausible explanation for the pattern you see in the statistics?

Sometimes, what seems to be a good story – how a particular initiative has produced a clear result – is in fact the product of changes in the way data are measured, collected or categorised. Data are particularly prone to change where the supplier itself has changed (e.g. from one contractor to another; or from a shift in organisational responsibilities; or a change programme within the supplier).

6. Is there any sense in which these data are obviously subject to confirmation bias – saying what those who commission it want to hear?

This is a difficult question for you to ask, because it could be that there is a tension between the role of statisticians as independent data collectors and your responsibilities for the successful delivery of policy. Nevertheless, it is an essential question to ask, even if the answer does not make for comfortable reading.

7. What level of assurance do you have over the administrative data?

The UK Statistics Authority has recently emphasised the importance of statisticians obtaining assurance over the underlying administrative data that

feeds into official statistics. So your statisticians ought to have a clear answer to the question about how they know the underlying data are reliable. If not you should ask them to get it.

8. How important are the data to the supplier – to meeting their own KPIs/success factors?

If the data supplier uses the underlying administrative data as the basis for a) its performance against a key performance indicator and b) for its returns to your statisticians, there is a potential for criticism of your statistics – because people may believe that the data supplier is incentivised to record the data in ways which make it look good. In this context, assurance is even more important.

9. What do you know about the processes by which the data have been compiled?

Like many of the other questions here, this is a remarkably simple question that can guide you as to how far you should rely on the statistics based on the data. If your teams don't seem to understand or trust the process, it's probably right to be concerned about the reliability of the official statistics.

10. How independent are the statisticians from the suppliers of the data?

The *Code of Practice for Official Statistics*¹⁴ (the *Code*) is the key document designed to ensure high quality statistical practice in government bodies. One of its key principles is the need for statisticians to act independently of both data suppliers and political decision-makers. So you should always ask your statisticians how they have complied with the independence requirements of the *Code*.

¹⁴ <http://www.statisticsauthority.gov.uk/assessment/code-of-practice/index.html>

Annex B: International approaches to audit arrangements and official statistics

Introduction

The National Statistician has prepared a guidance document¹⁵ around the use of administrative data. It describes the statistician's role in assuring and communicating the quality of administrative data for producing official statistics. It sets out the broader quality assurance context for determining the suitability of the administrative data, and emphasises the importance of: investigating the data source, establishing good relations and processes for the supply of the data, validation checks, and preparing clear explanations of the associated quality issues. It highlights the need to understand the audit arrangements of the administrative data within the critical questioning about the data source.

This report seeks to articulate the regulatory standard expected by the Authority on quality assurance and audit arrangements by, among other things, exploring international approaches. We reviewed some material published online from national statistical institutes (NSIs) and Eurostat to find examples of how audit features in their quality assurance approaches and how these are informed by the underlying quality frameworks.

ESS Quality Assurance Framework

Eurostat and its member states use the ESS Quality Assurance Framework¹⁶ (ESS QAF) with its six dimensions of quality: relevance, accuracy, timeliness and punctuality, accessibility and clarity, comparability, and coherence. Accuracy denotes the closeness of computations or estimates to the true values.

Eurostat emphasises the importance of providing quality reports to accompany official statistics, describing how each dimension is met. It has produced guidance¹⁷ to NSIs on the content of quality reports. It states that the accuracy section in quality reports should describe the methodology used, and identify the key variables and the main sources of random and systematic error of the statistics. An assessment of bias can be either described in quantitative or qualitative terms. It should set out the main sources of potential bias and reflect actions taken to reduce bias.

ESS QAF identifies some of the quality indicators that can be produced for administrative sources:

- Coverage, i.e. over- or under-coverage of the population that should have been included according to the definition or classification used

¹⁵ <https://gss.civilservice.gov.uk/wp-content/uploads/2012/12/Interim-Admin-Data-guidance.pdf>

¹⁶ <http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/introduction/>

¹⁷ http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/quality_reporting

- Errors in register variables – similar to measurement error in a survey (errors occurring during the collection)
- For event-reporting systems, an estimate of the rate of unreported events

ESS QAF recommends the preparation of a quality assurance plan in which statistical producers should:

- Assure the quality of data collection, including the use of administrative data
- Ensure the systematic examination of possible trade-offs within quality

These elements of the quality assurance plan should include the assessment of the collection arrangements within the data supplier organisation and encourage the proactive role of statisticians in influencing the system development to support the use of the data in producing official statistics.

ESS QAF identifies some specific activities that can be undertaken in relation to quality reporting to comply with Indicators in the ESS Code of Practice:

- Indicator 4.4 (the regular and thorough review of the key statistical outputs using external experts where appropriate) highlights the need for a plan for implementing quality reviews such as auditing and self-assessment for statistical outputs. This reference to auditing appears to be in relation to the conduct of peer reviews on the statistical outputs as opposed to the internal or external audit of the underlying data source. We found a number of examples of NSIs using the peer review approach for the auditing of statistical outputs, including Statistics Canada, Italy and Sweden. ONS has had a programme of National Statistics Quality Reviews that involved external experts and experts within the Government Statistical Service. The programme was recently reintroduced by the National Statistician, commencing with a review of the ONS Labour Force Survey¹⁸.
- ESS QAF also identified the need to carry out regular updating of business registers to ensure that the register reflects the changes in the population of businesses (under Indicator 7.3) and to conduct external evaluations of the methods used with external experts (under Indicator 7.7).
- In relation to administrative data the ESS QAF specifically highlights the distinction between statistical and administrative data processing and the use of appropriate validation rules and specific procedures for checking quality, in using the administrative data for statistical purposes (under Indicator 8.1).

Each of these elements of ESS QAF is relevant to the quality assurance of administrative data by statistical producers. However, these do not clearly set out the need for determining the assurance processes and robustness of measures taken by

¹⁸ <http://www.ons.gov.uk/ons/guide-method/method-quality/quality/quality-reviews/list-of-current-national-statistics-quality-reviews/nsqr-series--2--report-no--1/index.html>

data supplier organisations in the collection and processing of the administrative data.

Quality assurance arrangements in National Statistical Institutes (NSIs)

Our review of some material on the quality assurance approaches in various NSIs highlighted the importance of validation checks being conducted by statistical teams within the statistical producer organisations. The guidance appears in line with the ESS approach to reporting quality, identifying the steps to check and validate the data and statistics by the statistical producer. We found two NSIs – Statistics Netherlands and US Census Bureau – that specifically refer to the need for data supplier organisations to audit the administrative data and for the statistical producer bodies to assess the quality of the data identified as a result of the supplier's assurance activities.

Statistics Netherlands has developed a quality framework for administrative data sources¹⁹ for NSIs. It highlights three 'hyper-dimensions' of quality: Source, Metadata and Data.

- 'Source' reviews the quality aspects of the data source as a whole, the data collection organisation and the delivery of the data to the statistical producer. The detailed dimensions are: supplier, relevance, privacy and security, delivery and procedures
- 'Metadata' describes the quality aspects relating to the information required to understand and use the data: clarity, comparability, and data treatment by the data source keeper
- Data involves quality aspects that are mainly related to the accuracy of the administrative data such as over-coverage, under-coverage, unit non-response, item non-response and measurement – which includes activities such as 'external check: has an audit or parallel test been performed?'

It developed a checklist to assist NSIs in the assessment of the quality of the administrative data sources that covers the first two hyper-dimensions which both include some aspects of the assurance arrangements of data suppliers. The information is mainly captured through a structured study of data quality indicators under the 'Data' hyper-dimension.

The US Federal Committee on Statistical Methodology, the US Census Bureau, has also developed a tool²⁰ to assist in the assessment of the quality of administrative data. Its Data Quality Assessment Tool for Administrative Data provides a set of questions that can aid data providers in evaluating the quality of its administrative data, as well as assisting users (or the statistical producers) in determining the

¹⁹ <http://www.cbs.nl/NR/rdonlyres/0DBC2574-CDAE-4A6D-A68A-88458CF05FB2/0/200942x10pub.pdf>

²⁰ www.bls.gov/osmr/datatool.pdf

suitability of the data for an intended use. The tool comprises 43 questions to prompt the investigation of the data quality. It also uses a quality framework, with six dimensions: relevance, accessibility, coherence, interpretability, accuracy and institutional environment. 'Accuracy' has the same meaning as used in the ESS QAF. In relation to the assurance and audit of the underlying administrative data the US tool specifically asks:

- What investigations/analyses have been conducted that reveal data quality characteristics (such as Government Accountability Office reports, Office of Inspector General audits, internal agency reviews etc)
- Describe the checks the administrative agency performs to ensure the quality of the data and the typical results for your production processes
- Describe the principles, standards, or guidelines the agency uses as a basis to determine what is considered acceptable quality
- Describe the findings and corrective actions of studies, evaluations or audits to assess compliance with quality standards

The US tool is particularly helpful in highlighting the information that official statistical producers can obtain about the underpinning administrative data sources to better understand the quality implications for the statistical outputs. This information could be used in conjunction with the National Statistician's interim good practice guidance document on quality assuring and using administrative data (circulated to Heads of Profession for statistics in April 2014).

These tools have been influential among NSIs seeking to develop quality indicators for administrative data. In particular, Statistics Netherlands's tool has informed the development of a quality indicator instrument as part of the ESS network ('ESSnet') project on administrative data²¹ focusing on business and trade statistics. It is intended to be used by NSIs in routinely verifying the quality of administrative data received from data suppliers, and includes indicators for each of the ESS quality dimensions. The quality indicator instrument will be useful for statistical producers conducting their own quality assurance of the data received from suppliers, particularly in highlighting possible errors for further investigation.

The ESSnet project also identified checks to be conducted when first considering the use of an administrative data source for producing official statistics²². The indicators highlight the need to be clear about what is needed from the statistical use of the data and to build a comprehensive understanding of the administrative source. They suggest that producers compare definitions between the need and the source, and determine the coverage, such as geographical area, reporting delays, and thresholds

²¹ <http://essnet.admindata.eu/WorkPackage?objectId=4257>

²² <http://essnet.admindata.eu/WorkPackage?objectId=4252>

in size or quantity. Producers can also contact data suppliers and seek expert opinion from those managing the collection process, to become acquainted with practical aspects of the collection. The project also emphasises the benefits of comparing with other data sources, to gain a more objective view, particularly with a trustworthy source. It may be possible to use methods such as capture-recapture procedures to estimate duplicates in the administrative data. Also visual inspection will enable outliers and/or processing errors to be identified, through charting the data and making comparisons of distributions.

Findings

- ESS QAF provides some broad guidance on determining the quality of administrative data
- Activity in the national statistical institutes (NSIs) for member states has tended to focus on peer review of statistical outputs rather than checking the auditing and assurance arrangements within data supplier organisations
- Statistics Netherlands and the US Federal Committee on Statistical Methodology have developed tools to assist statistical producer bodies in determining the quality of administrative data
- The US tool is a helpful steer for official statistical producers when gathering relevant quality information from data suppliers about administrative data and forming a judgment of their suitability
- The ESS network project on administrative data for business and trade statistics has produced guidance for producers on making the initial judgment about the suitability of the data and a set of quality indicators that can be incorporated into routine checking of input administrative data

Wider lessons

The need to gather information about the auditing and assurance arrangements of administrative data suppliers has not been widely promoted among NSIs. However, some guidance is available for statistical producers to understand these arrangements and inform the use of the administrative data. The US quality assurance tool provides a useful starting point for investigating audit arrangements.

We have drawn on the audit items from the US tool, together with elements of good practice identified in the six case studies (Annex C) and from the ESS network project, to form a model describing the practice areas to consider when reviewing the quality assurance and audit arrangements for administrative data. It has informed a QA maturity matrix (see Part 4) which is designed to assist statistical producers in critically evaluating the robustness of the data for use in producing official statistics.

Annex C: Case studies illustrating the use of information about the audit of administrative data by producers of official statistics

C.1 ONS's Police Recorded Crime Statistics²³

Background to the data

Until 2012 police recorded crime (PRC) statistics were collated and published by the Home Office. In April 2012, responsibility for the production of these statistics moved to the Office for National Statistics (ONS). ONS also took over responsibility for the Crime Survey in England and Wales, with TNS-BMRB commissioned to conduct the survey. The Home Office retains responsibility for policy making in this area. It continues to collate and quality assure the PRC data prior to sending them to ONS for the production of the crime official statistics. There is a (unpublished) Memorandum of Understanding between ONS and the Home Office which outlines the roles and responsibilities of each department in the production of statistics about crime. ONS publishes a quarterly report which presents statistics from both PRC and the Crime Survey and it also publishes topic based reports throughout the year.

In January 2014, the UK Statistics Authority found that there was an accumulation of evidence to suggest that the underlying data on crimes recorded by the police may not be reliable and it removed the designation of National Statistics from the recorded crime statistics. It was retained by the Crime Survey for England and Wales.

ONS's recorded crime statistics are used by the police to monitor trends and by the Home Office to design and monitor policies and strategies. Other government departments use specific aspects of the statistics to monitor their own policies – for example Ministry of Justice uses analyses focusing on perceptions of crime and criminal justice to inform the development of criminal justice system reform policy. Police and Crime Commissioners use the statistics to make comparisons with similar forces and regional and national averages, and to monitor local targets. Local authorities use the statistics to gain a regional picture of crime rates and to monitor the impact of policy reforms on council services, service users and the broader community. The statistics are regularly used by ONS and the Home Office to answer Parliamentary Questions about crime. The statistics are widely reported in the media, reflecting the public interest in crime and the criminal justice system. Academics use the statistics as part of a wide range of criminal justice research and for teaching purposes. Voluntary organisations use them to assess the risks of victimisation across different groups.

PRC data are supplied to the Home Office every month by the 43 police forces in England and Wales, plus the British Transport Police. Data are submitted either via aggregated returns (Excel spreadsheets) or the new HO Data Hub which provides

²³ <http://www.ons.gov.uk/ons/re/crime-stats/crime-statistics/index.html>

record level data. The police are required to record crime in line with the National Crime Recording Standard (NCRS) and the Home Office Counting Rules for Recorded Crime (HOCR). All those who record crimes in police forces should be trained in the application of these standards.

Each police force has its own force crime registrar, who is responsible for overseeing compliance with HOCR and NCRS, and who is the final arbiter for whether a force should record a crime or make a 'no-crime' decision (to reverse the decision that an incident was a crime). Crimes can be reported in a variety of ways, for example to a call centre, to the police in person on the street, or by a third party and are recorded onto individual police force systems. The facts recorded about each incident allow the police to assess the matter reported to them and determine whether a crime has been committed.

Quality assurance

The Home Office Statistics Unit's Police Data Collection Section ensures that each force has submitted data and carries out basic quality assurance checks, such as comparing the current month's return with previous ones, identifying outliers or inconsistencies in the data. It raises queries with individual forces if there are revisions outside a certain tolerance level or if there are missing data or obvious errors found. The forces are then asked to check their data and resubmit them if necessary. The PRC data are then sent to the Home Office Crime Statistics team who carry out more validation checks and raise further queries with forces about data that appear inconsistent. A process map showing the stages of data quality assurance throughout the data cycle would be useful.

The Home Office supplies the PRC data to ONS who then carry out some further independent consistency checks, and examines the consistency with the crime survey data for equivalent offences to inform users about the relationship between the two series.

Audit

Police forces have internal audit procedures and the reports are generally discussed at each force's audit committee. After the launch of the NCRS in 2002, and up until 2007 the Audit Commission, commissioned by the Home Office, audited a sample of incident reports to check if crimes had been recorded correctly and published the results²⁴. For a number of years following completion of the Audit Commission's 2007 work, there was a lack of external scrutiny of recorded crime data.

Her Majesty's Inspectorate of Constabulary (HMIC) has the remit to carry out independent inspections of policing, including how crimes are recorded. Until recently HMIC had to be commissioned to carry out inspections, this is no longer the case. HMIC has carried out a number of recent reviews:

²⁴ <http://archive.audit-commission.gov.uk/auditcommission/nationalstudies/communitysafety/Pages/policedataquality0607.aspx.html>

- In 2012 HMIC published a review²⁵ of the quality of crime and incident data recorded by all 43 police forces in England and Wales plus the British Transport Police, and the arrangements in place to ensure that standards are maintained and improved. For the purpose of this review, a key finding was that HMIC ‘found limited evidence of forces directly assessing whether their own crime quality audits provided confidence that their crime figures gave an accurate account of their performance. Few forces compare crime audits with crime performance in any meaningful way’.
- In 2013 the Police and Crime Commissioner for Kent commissioned HMIC to conduct an inspection into crime recording in Kent Police. The report²⁶ concluded that ‘appreciably more needed to be done before the people of Kent could be confident that the crime and resolution figures published by the force were as accurate as they should be’.
- HMIC is currently carrying out an inspection of the integrity of crime reporting in England and Wales. It published an interim report²⁷ in April 2014 which highlighted some concerns with crime recording at the forces covered by the interim report. The full report is due to be published in autumn 2014.

Findings

- A lack of external scrutiny over a number of years for this high profile set of statistics has contributed to concerns about the underlying data, as noted in our earlier Monitoring Review ‘Overcoming Barriers to Trust in Crime Statistics’²⁸. The Public Administration Select Committee’s (PASC) recent report²⁹ welcomed HMIC’s decision to undertake a crime data integrity inspection in 2014.
- There is no clear outline of the process of data collection and quality assurance, in police forces, the Home Office or ONS. The statisticians could produce a process map to show the responsibilities of each party at each step in the process, and the stages of the quality assurance processes, identifying the areas of potential risk to the quality and accuracy of the data and the safeguards that are in place to minimise these risks. For example, a clear risk is at the start of the process if crimes are not recorded when they should be.
- Force Crime Registrars (FCRs) are of differing levels of seniority in police forces across England and Wales. PASC’s recent report noted that it is ‘essential that the Force Crime Registrar has not only had the requisite training but the necessary authority within the force to do their job’. FCRs should have clear backing from Chief Constables to record crimes with integrity.

²⁵ <http://www.hmic.gov.uk/media/review-police-crime-incident-reports-20120125.pdf>

²⁶ <http://www.hmic.gov.uk/publication/crime-recording-in-kent/>

²⁷ <http://www.hmic.gov.uk/publication/crime-recording-a-matter-of-fact-interim-report/>

²⁸ <http://www.statisticsauthority.gov.uk/reports---correspondence/reports/overcoming-barriers-to-trust-in-crime-statistics--england-and-wales.pdf>

²⁹ <http://www.publications.parliament.uk/pa/cm201314/cmselect/cmpubadm/760/760.pdf>

- ONS published a methodological paper³⁰ analysing the variation in crime trends between the Crime Survey for England and Wales and the police recorded crime data. This paper highlighted the growing divergence between the two sources since the cessation of external scrutiny in 2007.

Wider lessons

- High profile sets of statistics should be the subject of regular external scrutiny
 - Statisticians should make clear the level of scrutiny that they consider is necessary for each set of statistics
- Consider the roles of those who record, quality assure and sign off the data for publication
 - The statisticians should satisfy themselves that the data are managed by those who are sufficiently independent from reliance on targets or in the position of target setting, or whose performance will be judged on the basis of the data, within the data provider organisation
- Establish a Memorandum of Understanding or a Service Level Agreement between the statistical producer and data supplier bodies
 - This should clearly state the roles and responsibilities of those involved in the process
 - The key points from this agreement should be published
- Develop a clear process map
 - Explain each stage of data collection, processing and quality assurance, demonstrating who is responsible at each stage and the checks that they carry out on the data
 - This is especially important where data collections are complex and where several key stakeholders are involved in compiling the statistics.
- Identify the potential risks to the accuracy of the data and assess the safeguards that are in place to minimise the risk
- Carry out analysis of other key data sources which are used to corroborate the underlying data.

³⁰ <http://www.ons.gov.uk/ons/guide-method/method-quality/specific/crime-statistics-methodology/methodological-note--analysis-of-variation-in-crime-trends.pdf>

C.2 HSCIC's Social Services Activity³¹ and Expenditure³² Statistics

Background to the data

Community care is the process by which requests for social care help made to Councils with Adult Social Service Responsibilities (CASSRs) are translated, via assessment and care planning into appropriate services. Health and Social Care Information Centre (HSCIC) publishes a suite of statistical reports on adult social care in England. This includes the National Statistics, *Social Services Activity (Activity)*, which presents the number of referrals made to CASSRs, and the number of people receiving assessments, reviews and services funded by CASSRs. HSCIC also produces the National Statistics, *Personal Social Services Expenditure and Unit Costs (Expenditure)* which provides information about the money spent on adult social care by the social services departments. The underpinning information from councils is used in their day to day management of service users care plans and subsequent care packages.

HSCIC operates under the authority of the *Health and Social Care Act 2012*. This broadened the role of HSCIC, and established it as an Executive Non-Departmental Public Body. Following a social care data review, HSCIC is changing its social care data collections – to be implemented in 2014-15. These include a new Short and Long Term (SALT) Return.

HSCIC consults CASSRs and then confirms the arrangements for future collections in annual and quarterly letters to the Directors of ASSs. As part of the programme to replace the adult social collections, HSCIC circulates a monthly newsletter among councils to share ideas and examples of solutions through case studies from councils. It has seconded LA staff members to work on the implementation programme.

Care workers, social workers, care managers, and council administrative workers maintain operational databases used in the day to day management of service users' care plans. A care record is created for each service user and carers when they are assessed for social services. This is a record of the individual's needs and any services they require access to, this can also be referred to as a care plan. These records/care plans are maintained and updated when care packages for users are reviewed and changed.

³¹ <http://www.hscic.gov.uk/catalogue/PUB13148>

³² <http://www.hscic.gov.uk/searchcatalogue?productid=13760&topics=1%2fSocial+care%2fSocial+care+expenditure&sort=Most+recent&size=10&page=1#top>

Quality assurance

Councils routinely carry out data cleaning exercises on their systems to ensure the data are current (for example removing deceased clients). Councils use the data to monitor the quality of service delivery provided by their care teams.

Two secure data transfer systems are used in the adult social care returns – Data Depot and Omnibus. Omnibus has built-in validation routines to check the validity and completeness of the submitted data. A validation report for each CASSR highlighting any potential issues and the reports is emailed to the appropriate contacts. The expenditure information received via Data Depot also undergoes some system consistency checking, such as, identifying blank cells and auto-sums.

HSCIC's quality assurance activities include: system checks of missing data; investigations of missing information through contacting individual councils; examining internal consistency within and between tables; consistency checks over time and examining the plausibility of the data. For example, HSCIC compares the Omnibus data for the number of new service users to the number of service users receiving services and to the rate per population to check the plausibility of the data.

The results of HSCIC's checks are presented in an annex in *Activities* – it is a detailed presentation of the completeness of the main variables for each council and highlights any particular limitations with the data. It also explains if estimation was required to make up for missing information and the approaches taken.

Audit

HSCIC has not specifically documented the nature of audits conducted within LAs but it told us that it thought it might be possible to find out about such arrangements through its annual supplier questionnaire. Its Statement of Administrative Sources for these data highlights the routine quality assurance within councils but does not refer to their audit arrangements.

For the past two years HSCIC has prepared a review of the quality of the nationally submitted health and social care data. This comprehensive report³³ presents some examples of good practice by suppliers, as well as the steps taken within HSCIC to quality assure the data. It identifies some key messages about the importance of good quality data across the health and social care sectors and emphasises the difference between data validity and accuracy. It recognises the need to extend 'auditing of data quality beyond Payment by Results to other areas where poor data quality could impact most on direct clinical care'.

³³ <http://www.hscic.gov.uk/article/2021/Website-Search?productid=12280&q=quality+of+nationally+submitted+health+and+social+care+data&sort=Relevance&size=10&page=1&area=both#top>

There are a number of external sources of information about councils' social care data that can provide useful evidence to support judgments by statistical producers and users of the suitability of the data and factors affecting their quality, for example, such reports may highlight whether any concerns have been raised about the completeness or accuracy of the underlying information. The investigations may also indicate the overall effectiveness of the safeguards established by the local authorities in managing their services, and, in turn, provide some assurance or raise a concern about the quality of the administrative data. These sources include:

- information collated and published by the councils themselves about their systems and services through internal audits and local accounts of social service delivery
 - Health Quality Improvement Partnership (HQIP) provides support for providers to conduct social care audits³⁴ and has released the Quality Accounts Resource³⁵ to provide information for care providers planning their quality accounts submission
- information collated across councils:
 - by the Local Government Association (LGA) across councils about the social care services through its LGA Inform website³⁶ and
 - the Adult Social Care Outcomes Framework (ASCOF) indicators³⁷ collated by HSCIC on behalf of Department for Health
- the results of social care audits by and on behalf of regulatory bodies such as the Care Quality Commission (CQC)
 - CQC now regulates and inspects care providers against minimum standards of quality and safety³⁸
- national reviews e.g. by National Audit Office which published an overview report³⁹ on adult social care in England in March 2014

The findings from these reviews will not necessarily lead to quantitative estimates of quality but can provide a richer body of evidence to inform judgments about the suitability of the administrative data for use in producing and using official statistics.

Findings

- HSCIC has established detailed quality assurance processes to validate the data received from councils
- It provides clear and detailed guidance provided to councils in supplying the adult social data

³⁴ <http://www.hqip.org.uk/social-care-audit-frequently-asked-questions/>

³⁵ <http://www.hqip.org.uk/social-care-guidance-and-resources/>

³⁶ <http://www.local.gov.uk/about-lginform>

³⁷ <http://ascof.hscic.gov.uk/>

³⁸ <http://www.cqc.org.uk/content/fresh-start-regulation-and-inspection-adult-social-care>

³⁹ <http://www.nao.org.uk/report/adult-social-care-england-overview/>

- HSCIC's social care statistical team has some knowledge of the data issues associated with the data from their contacts with the suppliers. An explanation of specific issues identified during quality assurance is included in the quality sections of the *Activity* bulletin
- HSCIC's understanding of local issues was enhanced through the secondment of local government staff
- However HSCIC does not receive information about the steps taken by councils to audit their data and has not reviewed reports from NAO, local accounts or care quality accounts
- The statistical team told us that it could seek information from the councils about their audit arrangements using the questionnaire sent to all suppliers each year
- HSCIC also has quarterly and annual letters setting out the changes planned to collections and can use these to flag issues around audit and assurance
- HSCIC's annual quality report provides an important organisational statement about the issues that affect the operational uses of health and care information as well as the recording and of data underlying official statistics
- LGA Inform provides a useful information source for HSCIC's statistical team to learn about steps being taken by councils to improve social care delivery such as through peer review and local accounts
- HSCIC could follow up with some of the main organisations to find out whether data issues were identified during self-assessment and peer reviews and the efforts made to improve local accounts

Wider lessons

- Identify audit and assurance issues at an Institutional level:
 - An organisational level review of the quality of data collections provides important insight into the factors that affect the accuracy of administrative data for both the statistical producer and data supplying bodies, as well as for users – these can be issues that are of central concern to operational (e.g. clinical) uses, as well as for statistical needs
- Develop co-operative arrangements with data supplier organisation(s):
 - Establishing close ties with data supplier bodies is particularly important in developing a good understanding of the issues affecting data quality
 - The secondment of staff into the statistical producer body – and potentially from the statistical producer into the operational setting – is a valuable means of improving knowledge about the data collection and operational issues that may affect the quality of the statistics
- Determine the credibility of the data supplier organisation(s) to provide accurate and reliable data:

- Identifying the wider governance arrangements, such as through benchmarking schemes and quality accounts, will provide both statistical producers and users with a better appreciation of the issues affecting the accuracy of the statistics

C.3 HSCIC's Hospital Episode Statistics⁴⁰

Background to the data

Hospital Episode Statistics⁴¹ (HES) is a data warehouse held by the Health and Social Care Information Centre (HSCIC) which contains details of all admissions, outpatient appointments and Accident and Emergency attendances at NHS hospitals in England. The data are collected during a patient's time at hospital and are submitted to allow hospitals to be paid for the care they deliver. The HES system is designed to enable secondary use – that is, for non-clinical purposes – of the administrative data. Provisional HES statistics are produced and published by HSCIC on a monthly basis. The final annual reports are published as National Statistics.

The HES statistics have a wide range of users within HSCIC, by policymakers in the Department of Health (DH) and across a range of public and private health organisations. HSCIC also provides secure access to the underlying data to around 200 trained users from a range of organisations, such as government departments, hospital trusts and public health observatories. These users have access to pseudonymised⁴² record level data.

Quality assurance

Healthcare providers record patient data in a range of local patient administration systems, to support the care of the patient. The data are submitted to the Secondary Uses Service⁴³ (SUS) data warehouse. The raw data are then made available to commissioners and also copied into a database for processing. At pre-arranged dates during the year, data are extracted from SUS and then sent to HSCIC for processing and loading into the HES warehouse.

The HES data quality team in HSCIC validates and cleans the extract and derives new items. The team discusses data quality issues with the information leads in hospital trusts who are responsible for submitting data to SUS. The roles and responsibilities within HSCIC are clear for the purposes of data quality assurance, i.e. to assess the quality of data received against published standards and report the results of those assessments, but there is no central sign-off mechanism for the data submitted to SUS.

HSCIC has a well-developed data quality assurance process for the HES data, once the extract is received from SUS. It has about 700 data suppliers and uses an xml schema to ensure some standardisation of the data received. The use of the schema means that the dataset has to meet certain validation rules before it can be

⁴⁰ <http://www.hscic.gov.uk/hesdata>

⁴¹ <http://www.hscic.gov.uk/hes>

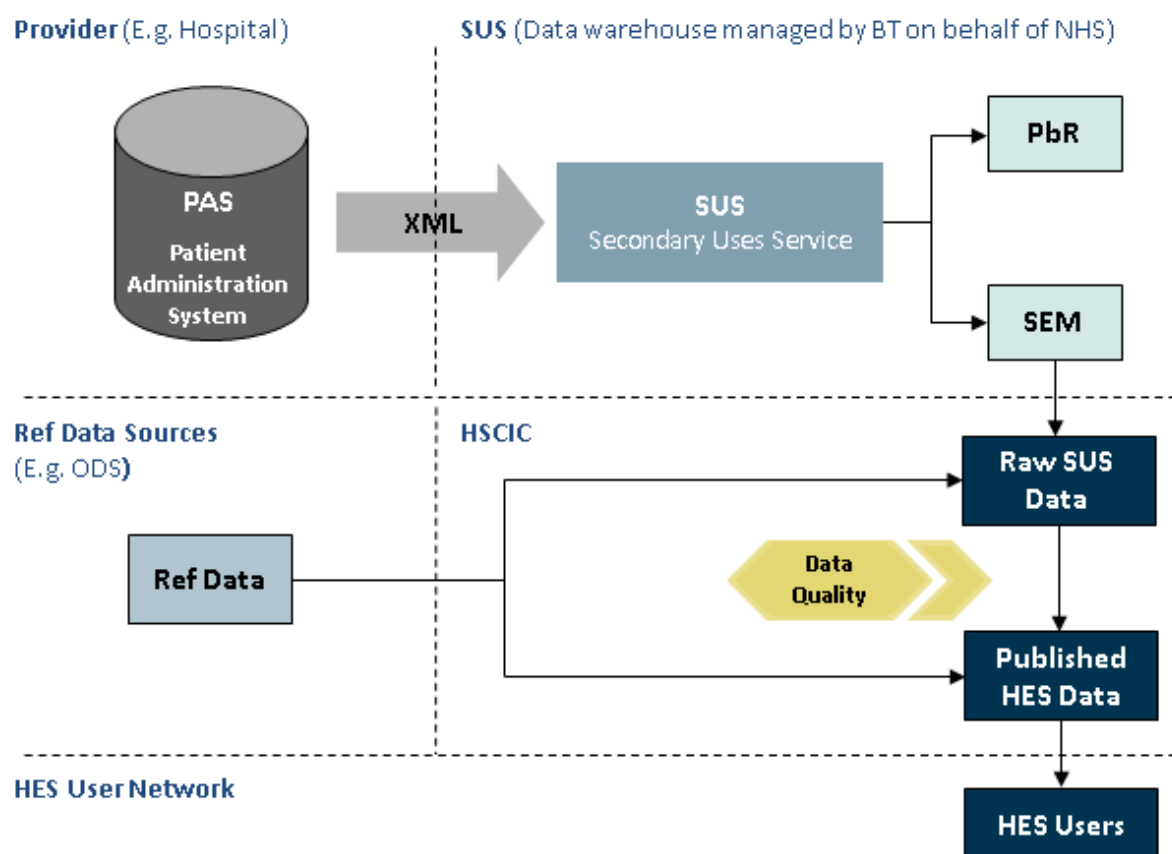
⁴² <http://www.hscic.gov.uk/dles>

⁴³ <http://www.hscic.gov.uk/sus>

submitted to SUS. HSCIC leads on the schema changes and consults the data suppliers about proposed changes.

HSCIC publishes detailed information about how the analysts collect and process the data used in the HES reports. This includes useful process charts to show the flow of data through the system (see Figure C3.1 below).

Figure C3.1: Illustration of the HES processing cycle



PbR = Payment by results extract

SEM= SUS extract mart

ODS = organisation data service (a website⁴⁴ with code reference library for health and social care organisations)

The range of guidance⁴⁵ also includes:

- Information about the data quality checks and data cleaning that the analysts carry out on the data after it is extracted from SUS, in order to prepare them for publication. HSCIC publishes the cleaning and derivation rules⁴⁶ for HES that have been developed over time

⁴⁴ <http://systems.hscic.gov.uk/data/ods>

⁴⁵ <http://www.hscic.gov.uk/media/1366/The-HES-processing-cycle-and-HES-data-quality/pdf/>

⁴⁶ <http://www.hscic.gov.uk/article/1825/The-processing-cycle-and-HES-data-quality>

- A further document⁴⁷ outlines the data quality checks performed on SUS and HES data when the Commissioning Data Set (CDS) is submitted via xml to SUS and must pass the validation to be accepted, the SUS Business rules and the checks carried out on the **Payment by Results (PbR)** dataset
- Data quality information for each year to date HES dataset is published in the monthly HES data quality report⁴⁸. These are published at the same time as the provisional year to date HES data. The statisticians can only check the validity and format of the data and not whether it is accurate, as accuracy checking requires a level of audit capacity and capability which the HSCIC does not currently possess
- HSCIC's second annual report *The Quality of Nationally Submitted Health and Social Care Data*⁴⁹ highlights issues around the recording of the underlying data that are used for HES, as well as examples of good and poor practice

Audit

HSCIC told us that its role is clearly defined in the *Health and Social Care Act 2012*⁵⁰ and that this doesn't extend to a regulatory role over the health care providers. For that HSCIC relies on the organisations that have such powers, such as, CQC or Monitor⁵¹.

HSCIC told us that the number of data suppliers (about 700) means that it is not possible to audit each individually. HSCIC said that the responsibility for the quality of the data submitted to SUS lies with the suppliers, even when held in the HES data warehouse.

DH currently contracts an independent external auditor to audit the accuracy of the data submitted to SUS, and used to calculate payment for activity, against that recorded in patients' notes (see section below). The HSCIC does not currently have the level of audit capacity and capability to do the same for data that is not used to calculate payment for activity, even though it may impact on the quality of care. However, HSCIC identified the importance of addressing coding issues in its annual quality report, *The Quality of Nationally Submitted Health and Social Care Data*. It noted that clinicians often see little direct value in their roles in using HES data and in ensuring good quality data, so it is often left to coders to interpret patient notes when coding. The quality report highlighted the concerns about quality of health records previously stated by Robert Francis QC in his report on the Mid Staffordshire

⁴⁷ http://www.hscic.gov.uk/media/13655/Data-quality-checks-performed-on-SUS-and-HES-data/pdf/Data_quality_checks_performed_on_SUS_and_HES_data.pdf

⁴⁸

<http://www.hscic.gov.uk/searchcatalogue?productid=14625&q=title%3a%22Provisional+Monthly+Hospital+Episode+Statistics%22&sort=Relevance&size=10&page=1#top>

⁴⁹ <http://www.hscic.gov.uk/catalogue/PUB11530/second-annu-data-qual-rep-2013.pdf>

⁵⁰ <http://www.legislation.gov.uk/ukpga/2012/7/contents/enacted>

⁵¹ <http://www.monitor.gov.uk/about-your-local-nhs-foundation-trust/regulatory-action/action-were-taking-nhs-foundation-trusts>

NHS Foundation Trust Public Inquiry⁵² and in Dame Fiona Caldicott's review of information governance⁵³.

From 2007 to 2012, the Audit Commission delivered a data assurance programme for PbR, the NHS tariff for paying acute hospitals. The work included an audit programme⁵⁴. The audits reviewed the key data that underpin payment in the NHS's PbR system. From 2013/14 the PbR data assurance framework is managed by the DH on behalf of NHS England and Monitor. DH commissioned Capita to deliver the PbR data assurance framework⁵⁵. In these audits Capita examined patient notes and compared these to what was recorded and submitted to SUS. While there are gaps in terms of what is included, for example, Clinical Commissioning Groups are to be given the option about what aspect of their data can be examined, it contains some information that could be usefully interrogated by the statistical producer teams as they consider the credibility of the data for statistical purposes.

Findings

- HSCIC publishes clear guidance for users about its data quality assurance processes on the patient level data in Secondary Uses Service (SUS), used to compile the HES statistics. This includes useful process maps
- HSCIC told us that it has clear internal processes for this data quality assurance work
- HSCIC has good knowledge of the quality of the data submitted to SUS for all fields where data standards exist for values and formats. However, aside from the information available via the PbR audits, HSCIC does not have knowledge of the accuracy of the data
- HSCIC is aware of wider audit work that could impact on some of the data presented in HES, but to date it has not used this as part of the quality assurance process
- HSCIC's annual data quality report provides examples of good and poor practice among health and social care data suppliers
- Overall HSCIC's quality assurance of data submitted to SUS for HES is thorough; however, its statistical teams could consider what other sources of information can be used to support their judgments about the suitability and accuracy of the data for producing official statistics

Wider lessons

- Gather information about audit arrangements from external data suppliers

⁵² <http://www.midstaffspublicinquiry.com/>

⁵³ <https://www.gov.uk/government/publications/the-information-governance-review>

⁵⁴ [http://www.chks.co.uk/Payment-by-Results-\(PbR\)-Assurance](http://www.chks.co.uk/Payment-by-Results-(PbR)-Assurance)

⁵⁵ <http://www.chks.co.uk/userfiles/files/PbR%20Key%20Findings%20Report%202013.pdf>

- The statisticians should consider other sources of existing audit that could provide them with some reassurance about the quality of the underlying data that they are using to compile the National Statistics
- Collate organisational information relating to the quality of underlying data in a central depository
 - HSCIC could create a central knowledge store to hold information about clinical audits across different health topics. This would allow the statisticians working on different topics easy access to these audits and less duplication of effort for each team to search for them. The statisticians could then consider, for each set of statistics, if any of the stored audits could be appropriate for their topic area or provide any insight into the underlying data.

C.4 ISD's Prescription Statistics in Scotland⁵⁶

Background to the data

The Information Services Division (ISD) a division of National Services Scotland, part of NHS Scotland, publishes a range of prescribing statistics. They cover different aspects of prescriptions for example dispenser remuneration, the number of items dispensed and costs of drugs.

Dispensing contractors, i.e. community pharmacists, dispensing doctors and appliance suppliers, are contracted by NHS Scotland to provide a service to the population of Scotland. To ensure drugs are available for dispensing when a patient arrives with a prescription, dispensing contractors buy prescription drugs in advance and then seek reimbursement for the drugs they dispense. NHS Scotland publishes statistics about its payments to dispensing contractors, these are remuneration for the service they provide and reimbursement for the products they dispense. These statistics are based on the data generated when a prescription is created, dispensed, and for which a claim is made.

As prescription expenditure covers in excess of £1 billion per annum and around 15% of the total NHS general revenue allocation in Scotland, there is strong interest from the Scottish Government and NHS service providers for information to ensure the clinical and cost effectiveness of these treatments. Prescribing data are also used for policy development, target monitoring, and for medical research such as clinical trials and epidemiology.

*Scottish Drug Tariff*⁵⁷ provides information about the prescribing, dispensing and reimbursement of medicines and appliances on primary care prescriptions. Types and value of dispensing fees are agreed with the Scottish Government and set annually. Details can be found in *Scottish Drug Tariff* and in Primary Care circulars⁵⁸ issued by the Government. Payments are derived from information gathered by Practitioner Services Division (PSD) in NHS Scotland, after the pricing of prescriptions has taken place. Reimbursement payments are made monthly to pharmacy contractors and will vary according to activity and claims. Some remuneration payments are set annually; some are now updated quarterly and some are set on a varying scale. Payments to dispensing contractors are made by PSD on behalf of the NHS Boards (the regional bodies responsible for the delivery of healthcare in Scotland). The data are collated and managed through a national payment system. The data warehouse brings together prescription, dispensed and patient information. Most prescription records (95%) generated by a GP in a GP practice have the patient's Community Health Index (CHI) number.

⁵⁶ <http://www.isdscotland.org/Health-Topics/Prescribing-and-Medicines/Community-Dispensing/Prescription-Cost-Analysis/>

⁵⁷ <http://www.isdscotland.org/Health-Topics/Prescribing-and-Medicines/Scottish-Drug-Tariff>

⁵⁸ [http://www.sehd.scot.nhs.uk/pca/PCA2013\(P\)21.pdf](http://www.sehd.scot.nhs.uk/pca/PCA2013(P)21.pdf)

The vast majority (about 92%) of prescriptions dispensed in the community are written by GPs, of which 99% are supported by an electronic prescription message. The remainder are written by other authorised prescribers such as pharmacists, nurses and dentists. When a GP writes a prescription, two forms are created; a physical form that the patient can take to a pharmacy and an electronic form (called an eMessage) that creates the first record of the payment cycle. When the patient then takes the physical prescription to a pharmacy, the pharmacist can access the eMessage and check the paper record against the eMessage for prescribing purposes. The dispensing contractors submit paper forms to PSD twice a month and PSD has ongoing access to the eMessages. The dispenser should also submit a dispensed e-message claim as well as supplying the paper copy of the prescription form for scanning. The prescription scanning process takes place on a monthly cycle.

PSD supplies information to the Prescribing Information System (PIS), which holds information on all NHS Scotland prescriptions dispensed within the community and claimed for payment by a pharmacy contractor (i.e. pharmacy, dispensing doctor or appliance supplier). The data include CHI numbers, prescriber and dispenser details, costs and drug information where available. Some research has estimated that these latter prescriptions account for around 6% of all prescriptions issued to patients. It is not possible to determine from payment data how much of the medicine dispensed to patients is actually taken in accordance with dosage instructions or why the medicine was prescribed.

The statisticians in ISD receive a dataset from PSD on a monthly basis. The data are stored in the Prescribing Information system (PIS), and ISD publishes quarterly official statistics about the payments made to dispensing contractors. ISD also run PRISMS, a web-based application which gives limited NHS Board access to prescribing information for prescriptions dispensed in the community from April 2004 onwards. The information is held centrally and the system is updated monthly. PRISMS can be interrogated to provide reports by individual prescriber, practice, locality, Community Health Partnership, NHS Board and for Scotland as whole. There are around 500 active NHS Board users of PRISMS and around 100 NHS Board users of PIS.

Quality assurance

PSD told us that there are approximately 5.3 million prescription claims per month covering 8.5 million items, supported by electronic messages for 4.3 million prescription messages for 6.5 million items respectively. The data go through several stages of checking on the payment system before being submitted to PIS for the statisticians to compile the statistics. These include checks of the pharmacy paper forms against statutory regulation, in-built validation checks of eMessages against the Dictionary of Medicines and Devices rules, and checks of what was dispensed compared with what was prescribed as recorded in the eMessage. There are also IT

validation steps in place for the loading of data into PIS; and checks of output by ISD when the data are first loaded (with tolerance levels for expected output compared to earlier time periods). In addition to this, the high use of the data increases the identification and amendment of anomalies in the data.

PSD has established business rules that determine item re-imbursement. If the electronic claim meets these business rules, then the claim can be automated. About 63% of item re-imbursements on the system are now automated, covering about 70% of GP prescribing. If the electronic claim fails to meet the business rules then the claim is sent for manual checks and item re-imbursement. Pharmacy contractors now have access to the data through online reporting and can also check their payments.

Routine monthly checks are carried out by PSD on a random sample of approximately 5% of prescription payments (the flow of prescription data and related checks are summarised in Figure C4.1 below). These check all data captured for payment and the accuracy of the payment calculation and have a target accuracy of 98% which is routinely met.

Audit

In addition to the routine financial audit described above, PSD told us that there is an annual, independent audit of the service. Also Community Pharmacy Scotland⁵⁹, the professional body representing community pharmacists, carries out further checks on the data. This is ongoing, and is generally on a random sample.

Audit Scotland has published some work in this area in 2003⁶⁰ and in 2013⁶¹. The latter report looked mainly at the financial implications of dispensing different types or makes of drugs, rather than auditing the payment process. While not specifically highlighting quality issues, it provides useful contextual information to support the interpretation of the statistics.

The statisticians at ISD told us that they do not generally explain the audit arrangements in the prescription statistical reports. They monitor quality issues by developing reports that can identify anomalies. They also share the data with NHS Boards who provide a further level of scrutiny.

Findings

- The prescribing data goes through a range of checks and balances but little information is published about this detailed assurance by PSD
- The statisticians also do not publish information about the financial checks that could be used to support the statistics

⁵⁹ <http://www.communitypharmacyscotland.org.uk/>

⁶⁰ http://www.audit-scotland.gov.uk/docs/health/2003/nr_030626_supporting_prescribing_km.pdf

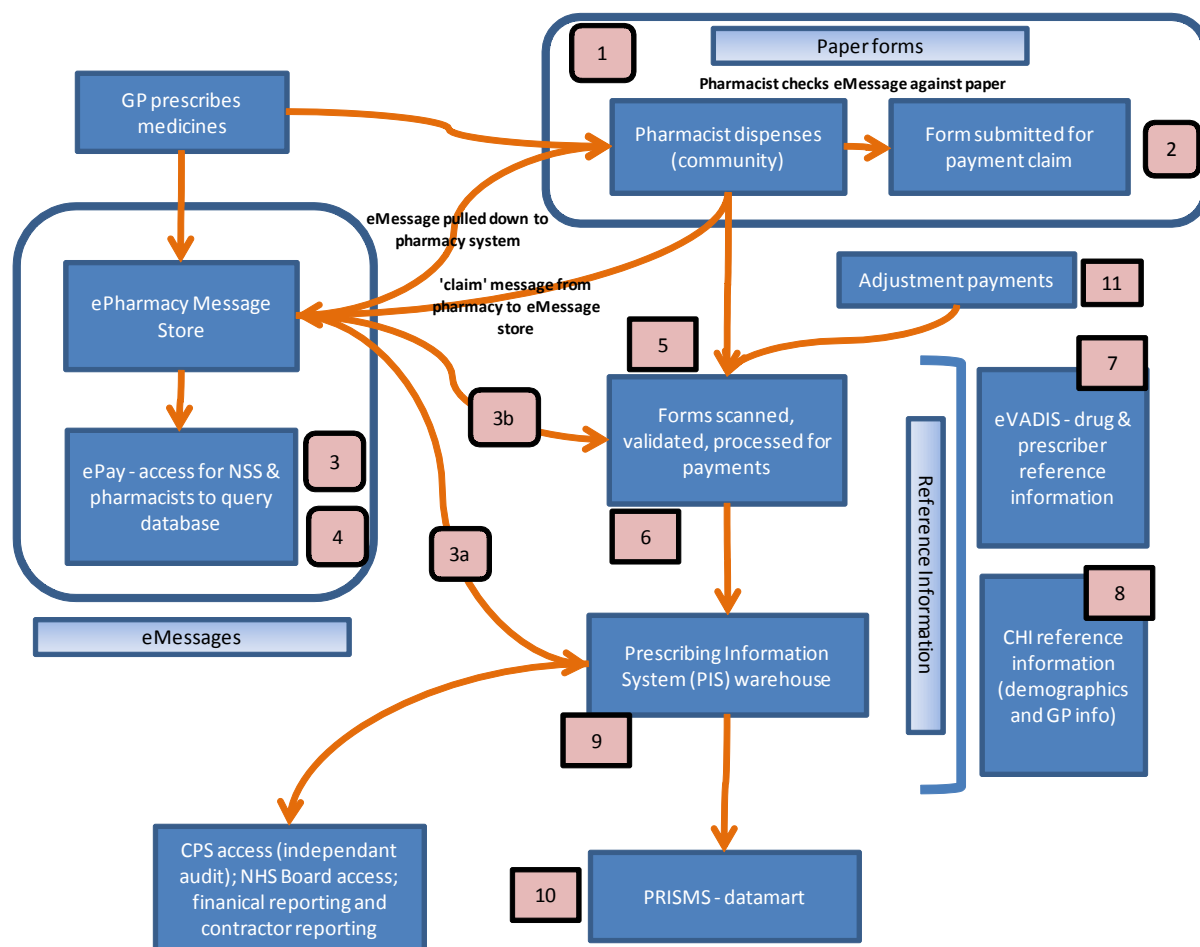
⁶¹ <http://www.audit-scotland.gov.uk/media/article.php?id=226>

- The statisticians could also look to include information about the independent financial audit that is carried out annually on these data

Wider lessons

- Fully document audit arrangements and identify the implications for the statistics:
 - Data that are used for payment purposes are likely to have detailed checks and quality assurance carried out on them, in addition to regular independent financial audit
 - The statisticians should look at these sources of audit and consider if they provide evidence of the robustness of the data and associated implications for the production and use of the statistics
- Gather information about audit arrangements from external data suppliers
 - The statisticians should obtain sight of these audit reports to satisfy themselves of:
 - a) the relevance of the audit for the statistics and/or
 - b) the quality of the underlying data
- Develop a clear process map:
 - Where the underlying data are subject to complex route through the process cycle, the statisticians should consider the inclusion of a process map to identify the stages in the data cycle and the data checks that occur at each stage

Figure C4.1: Process Map of ISD Prescription Statistics



Around 99% of prescriptions from a GP (on a GP10 form) produce an eMessage, this equates to 90-92% of ALL prescriptions. CHI is recorded on ~95% of prescriptions (allows patient details to be matched to prescription). Other prescription types: nurse; dispensing doctor; clinic; hospital; dentist; other (e.g. other nations).

Legend:

- 1 Guidance/statutory regulations for dispensing/endorsing prescriptions
- 2 Payment verification checks against guidance/statutory regulations
- 3 eMessages checked: Dictionary of Medicines & Devices (DM&D) rules; checks prescribed against dispensed; set tolerance levels e.g. same drug family; appropriate quantity.
- 3a If within tolerances then payment is automatic; 63% of all prescriptions are automated payment (70% of all GP prescriptions)
- 3b If outside tolerances eMessage is sent to keyer for manual checks.

- 4 Access to ePAY system for; NSS staff - manual queries and system checks; pharmacists - checks of own payments and dispensing
- 5 Validation rules: based on form types; dispenser type; item rules; business rules & tolerances
- 6 eMessages reconciled against scanned forms
- 7 eVADIS reference data: drug and prescriber reference files. Continually checked and maintained
- 8 CHI reference information: patient details, demographics and GP information. Continually checked and maintained
- 9 Warehouse brings together prescription, dispensed, and patient information; ~ 8.5 million items per month. 95% have CHI number (i.e. linked to a specific patient). Suite of reports run after each data load (monthly) for load checks; tolerances; exception reporting
- 10 Routine quality assurance and checks by analysts/statisticians of information and statistical products/publications produced from PRISMS and PIS
- 11 Adjustment payments made to correct over/under payments (Dec 13: 430 items out of ~ 9 million total number; value of £45k out of £8m total)

In addition to the checks and processes above, random sampling and external audit to assure NHS Board payments and ensure appropriate controls and checks are in place.

C.5 DWP's Work Programme Statistics⁶²

Background to the data

DWP's Work Programme is the government scheme in Great Britain to assist people who are long-term unemployed into sustained employment. The service is provided by employment support organisations through 40 contracts with 18 prime providers. These providers work with a larger number of sub-contractor organisations. Claimants are randomly assigned by the local Jobcentre Plus to a prime provider in their area.

The providers are paid when the claimants complete defined periods in work (usually after 6 months, but 3 months for those that are hard to place e.g. ex-prisoners) receiving the job outcome payment. The providers then can receive further monthly payments for each additional month in employment – i.e. sustainment payments.

DWP uses the statistics to monitor the performance of the programme. It has a business plan transparency indicator that it reports to HM Treasury every quarter. This metric is included in the official statistics bulletin. DWP monitors the performance of the providers through minimum performance levels which are also presented in the statistics bulletin.

Other main users of the statistics are: Parliament (through select committees), local government, employment support organisations (including the providers themselves), the third sector and the media. They use the statistics to hold the government to account, as well as to benchmark local delivery.

The data on referrals to the Work Programme are taken from Jobcentre Plus's administrative system (Labour Market System – LMS) used for administering customer claims and includes the claimants characteristics and claim details. The data on attachments to the Work Programme, as well as information about the payment of job outcome and sustainment payments, are submitted to DWP by the prime providers only (i.e. the sub-contractors do not supply data to DWP) for payment purposes. These records are submitted through the Provider Referral and payment System (PRaP) (see Figure C5.1).

Quality assurance

DWP statisticians undertook a large number of quality assurance checks during the early development of the Work Programme statistics to test: the reliability, completeness and level of disclosure of individual variables; the levels of duplicate, missing or contradictory information; and the consistency across computer systems and with management information. They also carried out a number of other investigations to determine the suitability of the data sources for use in producing

⁶² <https://www.gov.uk/government/collections/work-programme-statistics--2>

official statistics on the Work Programme. These included analysing the trends and variation in characteristic, time series and geographical breakdowns, as well as examining trends and differences in post-payment adjustment factors.

This quality assurance found no issues; it showed the data were robust, consistent with management information, and suitable for publishing. Individual variables were complete and consistent with existing sources, comparisons across systems were always within a 0.1% tolerance, including across key breakdowns and time series.

From this validation, the DWP statisticians then developed routine quality assurance procedures, including: data cleansing rules; checks on the scale of revisions as payment data are updated on the database, and automated checks against tolerances for a sample of tabulation tool tables. DWP reported⁶³ that individual variables remain complete and consistent with existing sources, comparisons across systems remain predominantly within a 0.1% tolerance, including across key breakdowns and time series.

Audit

a) Financial audit:

All Job Outcome payment claims are subject to an 'off-benefit check' before payment. This involves an automated check to match participant information on DWP's Customer Information System, to ensure that participants for whom Job Outcome payments are claimed are not claiming benefit.

The automated off-benefit check has a window of 15 days in which the check is applied, to allow for minor discrepancies between the details of the provider's claim and the details on Departmental systems. Job Outcome payment claims that fail this automated check are removed from the system (unless they can be validated) and not paid. Claims which pass the off-benefit check are released for payment, and are then subject to further post payment in-work checks.

Post-payment validation is performed every month to strengthen the controls against fraud and error in the Job Outcome payments reported to DWP by the Work Programme providers. This process involves selecting a sample of 33 claims per contract (for the latest 3 months statistics to be published) from the total population of Job Outcome payments that passed the automated off-benefit check and that were subsequently paid in that month. The sample is drawn at random from the population of Job Outcome payments paid in the sampling period and is of large enough size (across 3 months) to enable DWP to extrapolate error rates. DWP reviews each extrapolation round to ensure that it operated as intended.

⁶³ <https://www.gov.uk/government/publications/work-programme-official-statistics-background-information-note>

The sample is matched against HMRC P45 data to validate employment. Those that fail the HMRC check are validated by confirming employment with either the employer or the individual. Job Outcome payments that are found to be invalid are used to calculate the error rate which is extrapolated from the total population. The results of 3 rounds of validation (one for each month) are brought together every financial quarter to provide quarterly error rates used in the official statistics.

The primary purpose of the error rate is to extrapolate financial recoveries against all payments made to a contract in the extrapolation period based on the error rate, rather than for the sampled claims alone. Once the percentage of error has been calculated from the sample, the error rate is applied to the total paid to providers for the relevant three-month period, and the provider is then required to pay this back to the Department.

The error rates for the 40 contracts are used to derive adjustment factors which are then used to rate the official statistics to reflect final Job Outcome payments made to providers. The adjustment factor is derived using the number of the Job Outcomes which fail the post-payment validation process divided by the total number of Job Outcomes sampled. This ratio is applied to Job Outcomes (less the sample and those already validated) to adjust the official statistics.

Once the validation process has been completed, Work Programme providers have the opportunity to challenge its results. Time is allowed for providers to challenge and for DWP to assess and arbitrate any challenge. This process can take up to approximately 3 months, so that the official statistics on Job Outcomes for some providers may be revised slightly in the following quarterly release. DWP says that the affect of these revisions have so far been minimal.

The end to end post-payment validation process takes approximately 5 and 1/2 months to complete. The routine sampling, checks and production of error rates take just over 1 month and these are performed on the previous 3 months Job Outcomes payments.

b) External Audit

NAO conducted a data assurance audit⁶⁴ of DWP business plan transparency indicators and the data systems underpinning these metrics in 2012-13. NAO identified that DWP had undertaken a good assessment of the risks at the various stages of the collection process, including assessment of likelihood, impact and mitigating controls. It found that there are comprehensive desk instructions for each stage for the quality assurance and data validation processes; automated pre-payment checks on claims entered by providers including the 'off-benefit check' by DWP; and independent quality checks by a different team in DWP.

⁶⁴ <http://www.nao.org.uk/report/2012-13-review-of-the-data-systems-for-the-department-for-work-and-pensions/>

This audit concluded that:

- DWP has set out its strategy for the management and use of information, including its policies and standards to protect data and customer privacy, as well as how it re-uses information more effectively
- It has a data quality policy but needs to improve clarity over roles and responsibilities
- The Work Programme transparency indicator's system was rated as **'adequate but some improvements could be made'** – NAO highlighted that the latest version of the PRaP system had not been internally audited after its implementation.

DWP told NAO that it planned to do an audit of the PRaP system in the next financial year – it told us that it has completed its testing.

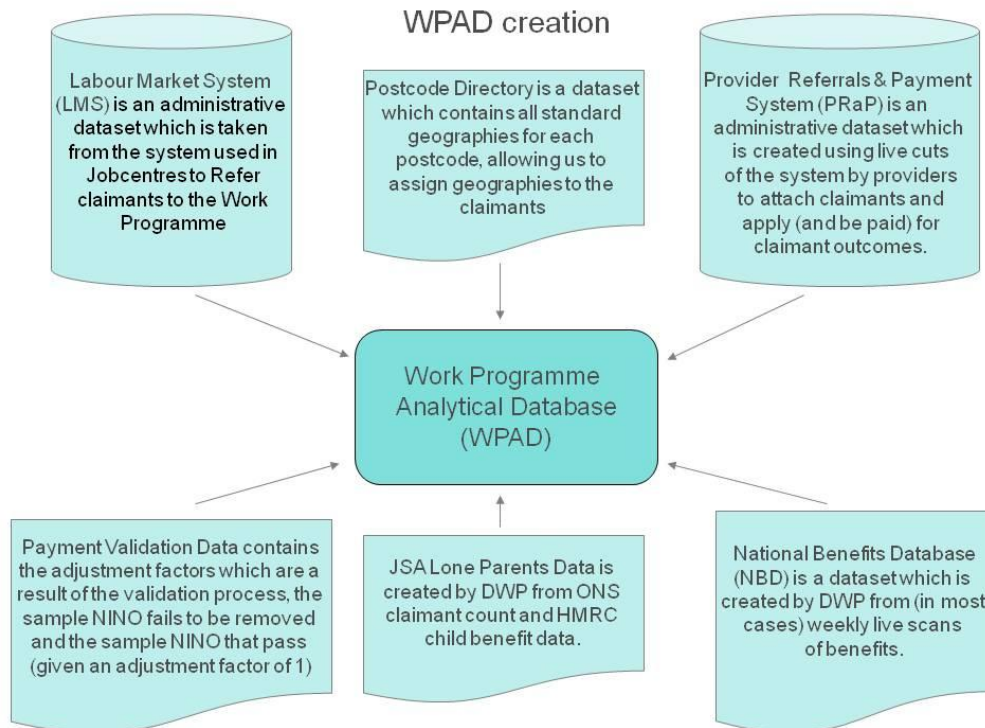
Findings

- The statistics are largely based on financial data received by the department for payment purposes
- Detailed validation and quality assurance both manually and through automated system checks
- Manual checks involve direct contact with claimants to verify benefit status
- NAO conducted audit of the data systems and rated them as **'adequate but some improvements could be made'** – recommending improvement in the clarity of the Data Quality Policy regarding roles and responsibilities and that DWP conducts an audit of the provider data system (PRaP)

Wider lessons

- Fully document audit arrangements and identify the implications for the statistics:
 - DWP has published information on its financial validation and quality assurance
 - It sets out the method to adjust the underlying data to prevent bias from delays to payment information undergoing validation and the impact on the statistics of operational delays in validation of payments
- Develop a process map:
 - DWP could provide further clarification of these arrangements such as by using the flow chart for the creation of the Work Programme Analytical Dataset, to indicate the main quality assurance steps and relevant quality indicators, e.g. PRaP error rate, % completeness of records, linkage rates
- Undertake internal audit of systems and processes for administrative data:
 - DWP provides reassurance of the robustness of the provider data by conducting its internal audit of the PRaP system

Figure C5.1: Diagram showing the various data sources feeding into the Work Programme Analytical Database



C.6 Health and Safety Executive's Injury Statistics⁶⁵

Background to the data

Administrative data on specified fatal and non-fatal injuries, occupational diseases and dangerous occurrences are collected under RIDDOR (*Reporting of Injuries, Diseases and Dangerous Occurrences Regulations 1995*). The regulation places a legal duty on employers to report certain workplace incidents to the relevant enforcing authority: HSE; local authorities; or, the Office of Rail Regulation (ORR).

HSE took over responsibility for the RIDDOR notification system in September 2011. Reporting by employers to HSE became predominantly online, using newly designed forms and online guidance. To produce the RIDDOR injury statistics, data are extracted each quarter from the RIDDOR database by the HSE statisticians, and a copy made for statistical purposes. HSE stores injuries information on an operational database for its area of enforcement. The database is maintained continuously. HSE combines data from its database with data provided by LAs and ORR to produce the RIDDOR injury statistics. These are National Statistics. The fatal injury figures are released quarterly, while the non-fatal injury figures are published annually.

Quality assurance

HSE makes a series of validation checks on the statistical data, looking into implausible data values, and making any necessary adjustments prior to publication.

In the case of fatal injuries, additional checks are made by correlating with additional sources of information, usually related to an investigation of the incident. All fatal injuries are investigated by HSE operational staff (the HSE inspectorate). A number of different sources of information are used in these investigations, just one of which is a formal RIDDOR notification from employers.

There are far more reports of non-fatal injuries than fatal injuries. A relatively small percentage of the major injuries are investigated further by the HSE Inspectorate; otherwise, there is no further detailed investigation of the events. The information provided by the employer when reporting an injury is taken largely at face value. HSE says that it has no cost-effective way to verify the information provided; however, it has commissioned a statistical audit of a sample of non-fatal injury records (see audit section below for further information) which provides some information on the scale of under-reporting.

The main quality assurance strategy involves a number of aspects of checking and validation; for example, conducting: system validation checks in the RIDDOR database to prevent incorrect data entry; data cleaning checks on data items in the statistical dataset; and consistency checks on the raw data, such as, comparisons

⁶⁵ <http://www.hse.gov.uk/statistics/causinj/index.htm>

against previous datasets. HSE also asks the relevant LA and ORR to confirm the accuracy of their injury data.

Audit

The regular review of injury records by the HSE inspectorate provides one type of audit of the injury data. Two other types of audit have been conducted by or on behalf of HSE: internal audit; and the statistical audit of non-injury data. HSE also compares the reporting of non-fatal injuries obtained through RIDDOR with the self-reported survey data from the Labour Force Survey (LFS). It commissions these survey questions to gain a view of work-related illness and workplace injury based on individual's perceptions and also presents these statistics in its annual statistical outputs.

a) Internal audit

In March 2012, HSE's Internal Audit team reviewed the RIDDOR system. This review was initiated by the statistics team following the transfer of responsibility of the injury notification collection system to the team in September 2011. The Internal Audit team's review focused in particular on how the process of reporting of fatal and major injuries was working and examined: wrongly allocated reports; backlogs of unallocated reports; the clarity of guidance documents in relation to reporting of incidents; the clarity of information provided in some aspects of the reports; and, the experience of local offices and HSE switchboard in responding to enquiries.

The team identified some areas requiring improvement. Following the audit, the statistical team:

- introduced some improvements to the online reporting form;
- provided guidance to assist the completion of the form;
- changed the review process to determine whether a incident was reportable or not; and,
- worked with the front line staff to understand better the potential impact of malicious reports on the injury statistics.

Internal Audit also conducted a follow up review in April 2013 to assess progress in addressing the required improvements and determined that appropriate actions had been taken.

b) Statistical audit of non-injury data

HSE commissioned a survey to check the information recorded on RIDDOR by speaking with the injured employees about the event. It enables HSE to better understand the issues that impact reporting on non-fatal injuries and the potential biases that occur as a result, as well as to provide information on the amount of time taken off work for reporting to Eurostat.

Based on the information obtained from interviewing around 2,000 injured people from a random sample of records of non-fatal injuries reported by employers, the survey found that:

- For injuries reported as major by employers:
 - 90% were confirmed as major
 - 10% were found not to reach the threshold for a major injury (that is, were over-reported)
- For injuries reported as over-7 day injuries by employers:
 - 60% were confirmed as over-7 day
 - 23% were under-reported and subsequently found to be major injuries
 - 17% were over-reported (i.e. were below the threshold required)

Overall, however, the survey concluded that the method was sufficiently rigorous to produce robust estimates of the average number of working days lost to workplace injury per worker to meet Eurostat's needs.

Findings

- Audit arrangements are integrated within RIDDOR process for fatal injuries through investigations conducted by the H&S inspectorate
- All notifications are reviewed by the H&S inspectorate, including non-fatal injuries
- Two reviews of the RIDDOR system were conducted by Internal Audit – the findings have been implemented by HSE
- A statistical sample audit has been conducted for HSE of non-fatal injuries – the results have not yet been published but has identified levels of under- and over-reporting
- HSE supplements data about non-fatal injuries by using self-reported data from the Labour Force Survey
- Some aspects of the audits have been published by HSE, and its quality assurance steps are explained alongside the statistics, but fuller information about the range of audit and investigation carried out on the injury data could provide further insight about the quality of the statistics to users
- HSE has developed a process map of its data supply arrangements
- HSE has not asked about audit arrangements in the LAs and ORR

Wider lessons

- Develop a process map relating quality assurance and audit arrangements into the operational and data supply arrangements:
 - HSE developed a process map to better help it understand any areas of weakness in the RIDDOR system (see Figure C6.1)
- Fully document audit arrangements and relate these to the implications for the statistics:

- HSE has documented the strengths and weaknesses of the statistics but did not specifically refer to the audit arrangements within these and their implications. Audit is not mentioned in the 'Background Quality Report' for the injury statistics
 - Publish a summary of the findings of the statistical sample audit and make clear the implications for the injury statistics in relation to use
- Gather information about audit arrangements from external data suppliers:
 - HSE obtained some information for the data it had collected but not for the data from other suppliers
- Use triangulation – identify alternative data and information sources e.g. from surveys – to verify accuracy of the data
 - HSE compares the administrative data on injuries with self-reported statistics from the Labour Force Survey

Figure C6.1: HSE process map for incidents reported by employers under RIDDOR (as at May 2014)

