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 guidelines[9] 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