

QUALITY MEASUREMENT AND REPORTING AT THE OFFICE FOR NATIONAL STATISTICS (UK)

Jane Birch

Office for National Statistics

Summary

Users of official statistics need access to a wide range of quality information so that they can judge whether an output is fit for their purposes. This paper describes how statistical quality is being addressed at the Office for National Statistics (ONS) in line with European standards.

The European Statistical System's six dimensions of quality, namely: relevance, accuracy, timeliness, accessibility and clarity, comparability and coherence; form the framework for statistical quality developments at the ONS. A set of quality measures and indicators has been developed to standardise quality measurement and reporting across survey outputs. This set will be implemented in statistical systems being developed as part of office modernisation.

Recent and ongoing developments at the ONS will be outlined, including the identification of Key Quality Measures, developing quality indicators for administrative data, and finding ways to report quality to general users.

1 INTRODUCTION

This paper gives an overview of how statistical quality is being addressed at the UK Office for National Statistics (ONS). The ONS approach is to standardise and systematise provision of quality information across outputs in line with European standards. A standardised approach to quality enables users to not only assess the intrinsic quality of individual outputs in terms of their strengths and limitations, but also to make quality comparisons between outputs.

This paper will describe:

- how quality is being standardised at the ONS (and ultimately across other government departments within the UK Government Statistical Service [GSS]) in line with European standards;
- the development of quality measures and Key Quality Measures;
- ongoing work on quality indicators for administrative data; and
- ongoing work on finding ways to report output quality to general users who lack statistical knowledge.

1.1 The European Framework: The Six Dimensions of Quality

The ONS has based its approach to statistical quality on the European Statistical System's (ESS) six dimensions of quality, namely: relevance, accuracy, timeliness and punctuality, accessibility and clarity, comparability and coherence. Definitions for these dimensions can be found in the document 'Definition of Quality in Statistics' (Eurostat 2003). These six dimensions encapsulate a broad understanding of quality, offering insight into the diverse needs for output quality of different users.

2 THE GUIDELINES FOR MEASURING STATISTICAL QUALITY

The Quality Measurement and Reporting (QM&R) team at the ONS has used the six ESS quality dimensions as a framework to develop a set of quality measures and quality indicators. These are published in the 'Guidelines for Measuring Statistical Quality' (ONS 2004). In addition the Guidelines draw upon advice in the Standard Quality Report (Eurostat 2002) on specific information to cover when measuring and reporting quality.

The aim of the Guidelines is to outline best practice for measuring and reporting statistical quality at all stages in the output production process. It is not intended that output compilers report quality against

all of the quality measures and indicators contained within the Guidelines: many of the measures will not be appropriate for the type of output at hand. Instead, it is intended that output compilers consult the Guidelines to identify those quality items that ensure the thorough reporting of statistical quality to users, so that users can assess whether a statistical output is of sufficient quality for their needs.

The Guidelines contain 130 quality measures and indicators, most of which are qualitative and require textual descriptions. Table 1 gives examples of some of the qualitative indicators from the Guidelines.

Table 1: Examples of qualitative indicators from the Guidelines for Measuring Statistical Quality

Qualitative indicator		ESS Dimension
1	Describe and classify key users of output	Relevance
2	What type of sample design was used?	Accuracy
3	Frequency of publication.	Timeliness.
4	Are there any restrictions placed on the use of this dataset or output, eg any contractual agreements limiting use?	Accessibility
5	Describe and assess impact on the data of any changes in concepts, definitions, classifications and methods over time.	Comparability
6	Compare estimates from other sources, and include any known reasons for the differences.	Coherence

Other quality measures or indicators are quantitative. Examples of quantitative indicators are shown in table 2.

Table 2: Examples of quantitative indicators from the Guidelines for Measuring Statistical Quality

Quantitative indicator		ESS Dimension
1	Standard error estimates (or coefficients of variation) for key estimates.	Accuracy
2	Estimates of rates of undercoverage, duplication, ineligibility and misclassification.	Accuracy
3	Scanning and keying error rates.	Accuracy
4	Coding error rates.	Accuracy

The majority of the quantitative quality indicators relate to the dimension of accuracy. Quantitative indicators for all dimensions have not been identified or developed, although there is scope for considering ways to quantify the comparability of outputs (see for example the scoring system to measure the comparability of data across regions developed by Linden et al 2004). In addition, simple measures such as 'Number of web hits' give a numerical indication of the relevance of an output for users' needs, and an indication of its accessibility. Punctuality can be simply demonstrated by counting (in days) any time lags between the scheduled and actual release dates.

2.1 Quality measures and indicators

The Guidelines contain many items that are quality **indicators**, rather than quality **measures**. It is rare that the quality of an output can be directly measured. For example, it is often not possible to measure non-response bias as the characteristics of non-responders can be difficult, and costly, to ascertain. In those instances where direct quality measures are not available, quality indicators can be used to give an insight into output quality. In the case of non-response bias, in practice often the response rate is used as a proxy for non-response bias. This is not an ideal situation, but quality indicators are sometimes the best available information on output quality.

2.2 Structure of the Guidelines

In order to assist output compilers when using the Guidelines, the document is structured in two ways:

- according to the order in which statistical outputs are produced, from the initial design stage, through data collection, processing and analysis until finally the output is disseminated. This enables users of the Guidelines to find the relevant quality information that they can produce for each stage; and
- each quality measure or indicator is given a reference to the relevant quality dimension – for example, ‘editing rate’ and ‘standard error’ fall within the dimension of ‘accuracy’, while the quality indicator ‘Time lag: reference date to release of provisional output’ falls within the dimension of Timeliness.

In addition, to ensure the standardised application of the Guidelines across the office and the GSS, supplementary information is provided. This supplementary information consists of:

- examples of how to populate an indicator (usually real-life examples from published outputs);
- formulae for quantitative measures;
- the types of output for which each measure is applicable (eg time series, business surveys); and
- the frequency of their production (eg in the event of a discontinuity, for each survey cycle etc).

2.3 Implementation of quality measures and indicators

Another means to ensure standardised delivery of quality information within the ONS is the implementation of quality measures as part of office modernisation. The ONS is currently undergoing a major process of standardisation of processes and tools used for output production. This overhaul of processes and systems is referred to as the Statistical Modernisation Programme (SMP). The aims of the Programme are to standardise the way outputs are produced, and to provide a common set of software tools to carry out functions that are generic to many business areas.

The quality measures and indicators developed for the Guidelines are ultimately to be implemented via the software tools being developed by projects within the SMP. The majority of the measures within the Guidelines are qualitative, and these are to be implemented within the metadata repository currently being developed. The aim is to generate text fields within the metadata system, which will be manually or automatically populated with qualitative quality information. Quantitative indicators from the Guidelines will be implemented via other statistical tools being developed by the SMP; for example variance estimates for most surveys will be calculated by the standard estimation tool (the Generalised Estimation System developed in Statistics Canada). External users will be able to access the quality information associated with each output via the National Statistics website (which will interface with the metadata repository, where all the quality measures and indicators are to be stored).

The implementation of quality measures within the SMP is at a fairly early stage: provisionally it is anticipated that quality measures will be systematised in standard office software tools by 2006.

2.4 Extension of the Guidelines for Measuring Statistical Quality

At present, the Guidelines have been developed with a focus on quality reporting needs for survey outputs. Over time quality measures and indicators for non-survey outputs will be incorporated into the Guidelines. In addition, there are some areas of methodology that have not been covered, for example suitable quality indicators for index numbers. Again the Guidelines will evolve to include recommended quality information for those areas of methodology currently excluded.

3 KEY QUALITY MEASURES

A further development arising out of the quality measures developed for the Guidelines is the identification of 11 Key Quality Measures (KQMs). These have a dual function: to provide a useful summary of overall output quality and to enable performance monitoring across the office (ie to act as statistical performance indicators). The definition of KQMs is:

The smallest set of important, informative, achievable quality measures and indicators which provide a summary of the overall quality of an output.

The KQMs were chosen from the array of quality measures in the Guidelines, according to the following selection criteria:

- the need for output and process quality indicators;
- the indicators must be achievable;
- whether there is a 'gain in confidence' from the indicators (ie different values of the indicator are related to a loss or gain in confidence in the quality of an output);
- whether they can fulfil a 'real-time analysis' function (ie where an indicator can have different values throughout the output production process); and
- ensuring coverage of the ESS quality dimensions.

The 11 KQMs are shown in table 3. These 11 KQMs provide a minimal set of compulsory quality information that should be provided for every statistical output, where they are appropriate. It was stated earlier that the Guidelines represent best practice and as such they contain a long list of suggested quality measures for measuring and reporting quality. By identifying 11 compulsory KQMs, users will have access to the minimal amount of quality information required for an overall view of quality, without overburdening business areas with demands that they provide unrealistic amounts of quality information.

Table 3: Key Quality Measures

Key Quality Measures		ESS Dimension
1	Standard error estimates for key estimates	Accuracy
2	Overall unit response rate by sub-groups	Accuracy
3	Total contribution to key estimates from imputed values	Accuracy
4	Key item response rates	Accuracy
5	Editing rate (for key items)	Accuracy
6	For estimates revised on a regular basis: estimate the likely revision between provisional and final estimates and include known reasons for differences	Coherence
7	Compare estimates from other sources, and include any known reasons for differences	Coherence
8	Describe uses supported and where possible how the data relate to needs of users	Relevance
9	Identify known gaps between key user needs, in terms of coverage and detail, and current data	Relevance
10	Provide a statement of the national / international agreed definitions and standards used	Comparability
11	Time lag: reference period to release of provisional / final output	Timeliness

Not all of the KQMs fulfil all of the selection criteria above, but the whole set contains both output and process indicators, is achievable, contains some measures that represent a 'gain in confidence', has some real-time information and covers five of the six ESS dimensions. There are no quality measures within the Guidelines for the dimension Accessibility and Clarity that would be considered KQMs.

At present, they display a bias towards the quality reporting needs of survey outputs (which is to be expected, since they originate from the Guidelines, which also has a survey-bias). It is anticipated that the list will be expanded upon in the future to take account of quality reporting requirements for non-survey outputs.

3.1 Using Key Quality Measures for management purposes

As well as being used to provide a summary of overall output quality, the KQMs can be used in performance monitoring for the purposes of continuous quality improvement and to measure improvements in re-engineered outputs (for example, Labour Market Statistics [LMS] are currently undergoing re-engineering in order to streamline their processes and modernise systems used to produce LMS outputs). To be used as performance gains measures, KQMs need to be compared before and after re-engineering to assess the extent of improvements afforded by re-engineering. In

their performance monitoring function, the KQMs fulfil the office requirement for statistical key performance indicators: other non-statistical key performance indicators may include personnel or financial information, which are not covered by the KQMs.

Dynamic KQMs (ie those that provide real-time information) can be used for management purposes, for example 'response rate'. At present it may be the case that response rates are calculated only once at the end of the data collection stage. However, it is possible for daily response rates to be calculated, providing useful management information on where response rates are falling below target and where remedial action may be required.

It is envisaged that, in the future, the KQMs will be provided to internal managers via a Survey Management Information System, and functionality for dynamic KQMs will be identified in or built into current or future systems with capacity to re-calculate these quality measures on a regular basis.

4 ONGOING AND FUTURE WORK

This section outlines some of the current and future work plans for the QM&R team. These are in addition to the ongoing work on extending the Guidelines for areas of methodology currently excluded, and co-ordination across several Modernisation projects to ensure that quality measures are implemented within systems.

4.1 Quality indicators for administrative data

The aim of this work is to identify suitable information for assessing the quality of administrative data that are used to compile statistical outputs (Thomas, 2003). Administrative data are primarily collected for official, non-statistical purposes (eg tax records, medical records or vehicle registrations). However, administrative data can also be used for statistical purposes, offering the opportunity to increase available data while reducing the compliance costs that would be incurred were this information to come from specially designed surveys. In addition, administrative data are often collected for whole populations, affording great advantages in terms of coverage over sample survey data.

However, one of the challenges in using administrative data is that they are produced by external providers; hence statistical institutions have less control over quality. This means that the ONS needs to find ways to assess the quality of administrative data. The first step towards assessing quality is to set up a single gateway for all incoming administrative sources, where they can be processed in a standard way. Quality assessment is likely to be a part of the processing of incoming administrative data in the future. This will require a set of quality indicators to audit whether the incoming data are fit for purpose. The Guidelines do not currently provide suitable quality indicators for administrative data, and therefore the QM&R team is devising guidelines for the assessment of administrative data that are used for statistical purposes.

4.1.2 Approach to assessing administrative data quality

The approach to assessing quality of administrative data is two-fold: firstly to adopt the Eurostat (2003) recommendation that quality reports are produced for each administrative data source, as well as for each statistical output in which administrative data are used; and secondly to use the ESS dimensions of quality as a framework to identify the measures and indicators required to assess the quality of administrative data. For the first strand of this work, it is recommended that the source-specific report for each administrative data source is available to internal users only, while the statistical output report is the standard quality report that is made available to external users. The quality measures and indicators to be used for administrative data are therefore divided into those that are used to assess the quality of the source, and those that are used to assess the quality of the statistical product which has drawn upon administrative data.

Examples of source-specific and product-specific quality measures and indicators for administrative data are shown in Table 4.

Table 4: Examples of source-specific and product-specific quality measures and indicators for administrative data

	Source-specific quality measure or indicator	Product-specific quality measure or indicator	ESS Dimension
1	Describe the envisaged uses of the data and / or all statistical products known to require inputs from the source.	Describe each administrative source used as an input in the statistical process.	Relevance
2	Describe all coverage problems.	Assess the effects of any coverage problems on the statistical product.	Accuracy
3	Describe the timescale since the last update of data from the administrative organisation.	Assess the effect of any lack of timeliness on the statistical product.	Timeliness and punctuality
4	Describe in full any constraints on the availability of data at the required level of detail.	Assess the effects of any constraints on the availability of data on the statistical product.	Accessibility and clarity
5	Describe any changes in concepts, definitions, data collection purposes, data collection methods, file structure and format.	Assess the implications of changes in concepts, definitions, data collection purposes, file structure and format for the statistical product.	Comparability
6	Describe the common identifiers of population units.	Assess the effectiveness of record matching techniques.	Coherence

It should be stressed that this work is ongoing, and the quality measures and indicators listed above are still undergoing consultation and development. It is anticipated that the quality measures and indicators to be recommended for administrative data will be incorporated into the Guidelines by 2006.

4.2 Reporting quality to general users

Not all users of official statistics have sufficient statistical knowledge to understand how statistics are produced and the various elements that impact upon output quality, such as sampling and non-sampling error. The general user may refer to official statistical outputs for figures upon which to base allocation of funding, for data to inform policy development and policy decisions, or they may require statistical data to inform business planning. These users may take official statistics at face value, presuming that all estimates produced within an output are of equally good quality. Therefore the QM&R team are investigating ways to communicate quality to the general user in terms that can be understood by those with little or no statistical knowledge. The aim of this work is to ensure that general users are advised of which data are of good quality and where data should be used with caution. It is hoped that, armed with this information, the general user will be able to assess whether output quality is sufficient for their needs. This work is in its initial stages, and some of the ideas currently being considered to report quality to general users are outlined below. All of these are provisional, as the QM&R team are in the process of deciding which should be investigated further.

4.2.1 Coefficient of Variation grading system

The QM&R team are currently reviewing whether quality grading systems based on the size of the coefficient of variation (CV) would be a feasible and reliable way to inform general users of the precision of estimates of level. Such grading systems are used elsewhere in the GSS and in other National Statistical Institutes, using simple pictorial means (such as ticks) or letters (eg A – E) to communicate various categories of precision of estimates. Each category of precision is denoted with a symbol or letter, which is defined in terms such as ‘excellent’, ‘acceptable’, ‘use with caution’ or ‘unreliable’ (see for example Statistics Canada’s use of a CV quality grading system on the Unified Enterprises Survey).

It is not yet known whether a CV-based quality grading system would be feasible across all ONS outputs, for similar types of survey, or whether they should be set for each individual survey. In addition, there is the issue that the CV only indicates one aspect of quality to users, namely the precision of the estimate. However, general users are currently provided with little accessible information on quality, and the CV-based quality grading systems have been used with success elsewhere.

4.2.2 Imputation indicator

For the 2001 Census, a simple imputation indicator scheme was used to signify the proportion of imputed people in an area. This used the following grading system:

1	Less than 5%
2	5% and less than 10%
3	10% and less than 20%
4	20% and over

This imputation indicator scheme is being reviewed to assess whether it would be a suitable means to communicate quality for other types of output. This scheme informs users of the amount of 'real' as opposed to imputed data that are used to compile an output. However, there is no simple equation between various levels of imputation and quality in that the proportion of imputed data tells users nothing about the quality of the imputation itself.

4.2.3 Usability Statements

A simple way to communicate quality to general users is to devise Usability Statements. These statements may contain advice on:

- how statistics are used, eg across six to ten key applications of the data;
- common pitfalls to avoid when using the data, eg not over-stretching data to provide information for small sub-groups;
- the strengths and limitations of different data sources, such as different sources of information on unemployment;
- summaries of conceptual issues;
- summaries of accuracy and precision; and
- aspects of the output production process which may have had a significant impact upon the quality of the output, for example where the sampling design or data collection methodology have implications for output quality.

The content of these statements would be specified in consultation with output providers and experts within the Methodology Group (MG) at the ONS. It is envisaged that Usability Statements would be a compulsory addition to the KQMs, and that examples on how to populate the statements would be devised in a style that is suitable for a statistically unsophisticated audience.

4.2.4 Other ways to communicate quality to general users

Other ideas currently being explored by the QM&R team, but which are less well-defined, include a quality grading system for estimates of change and an overall 'index' of quality.

5 CONCLUSION

The ONS has made considerable progress in specifying quality measures and indicators for survey outputs in line with European standards. The quality measures contained in the Guidelines for Measuring Statistical Quality are to be implemented in systems being developed as part of office modernisation. In addition, with the identification of Key Quality Measures, the office has addressed the need for a minimum set of quality information that output providers must produce where appropriate, so that users have an overall summary of output quality.

Ongoing work includes extending the Guidelines to address the quality measurement and reporting needs for non-survey outputs, such as administrative data. In addition, simple means to communicate quality to statistically unsophisticated users are currently being explored.

For more information on any of the issues discussed in this paper please contact the Quality Measurement and Reporting Team at:

Quality.measurement@ons.gov.uk

6 REFERENCES

Eurostat (2002): *Standard Quality Report*, Luxembourg.

Eurostat (2003): *Methodological Documents – Definition of Quality in Statistics*, Luxembourg.

Eurostat (2003): *Quality Assessment of Administrative Data for Statistical Purposes*.

Linden, H., M. Petrakos and P. Stavropoulos (2004): *Use of Proximity Data for the Assessment of Geographical Comparability*, Germany.

ONS (2004): *Guidelines for Measuring Statistical Quality*, [online]. UK. Available from: http://www.statistics.gov.uk/methods_quality/publications.asp [Accessed 24 June 2004].

ONS: *Census Quality Indicators for Local Areas, London* [online]. UK. Available from: <http://www.statistics.gov.uk/census2001/methodology.asp> [Accessed 24 June 2004].

Statistics Canada: *Unified Enterprise Survey Information Package Contents* [online]. Canada. Available from: http://collection.nlc-bnc.ca/100/200/301/statcan/unified_enterprise-e/html/ues5.htm [Accessed 2 July 2004].

Thomas, M. (2003). *Guidelines for Assessing the Quality of Administrative Data for Statistical Purposes (Second Draft)*, London: ONS. [Internal paper available upon request from marilyn.thomas@ons.gov.uk].