REPORTING ON DATA QUALITY AND PROCESS QUALITY

Joy Dobbs, Catherine Gibbins, Jean Martin, Pam Davies and Tricia Dodd, Office for National Statistics
Joy Dobbs, ONS, Social Survey Division, 1 Drummond Gate, London, SW1V 2QQ

Key Words: quality surveys reporting data process

Summary

This paper considers the development of definitions of quality over time, from the traditional measures of components of total survey error to wider definitions that encompass quality indicators and process measures. It also describes the way in which the focus has shifted so that producers now have a greater responsibility to report on quality for the user. A typology is developed looking at the reporting of product and process quality and the extent to which their effects can be quantified. Examples of three ONS initiatives on reporting quality are described.

1. Definitions of quality

1.1 Traditional approach to survey quality

National Statistical Institutions (NSIs) and other survey organisations have always placed great emphasis on high quality data and methods. Over time, the focus has widened to cover the quality of their whole range of outputs and services: not only the statistical data and information produced but also survey advice/consultancy and the service provided to users and customers throughout a survey project. This paper, however, focuses on outputs only, concentrating on reporting on the quality of the survey data and estimates.

The traditional approach to measuring and reporting survey quality has been based on the concept of total survey error, with quality defined as the absence of error. Total survey error can be classified or modelled in different ways. Groves (1991) divides errors into errors of non-observation and errors of observation, and then subdivides each by the different sources of error. A further category, processing errors, is often added to cover errors introduced during processing which were not present when the data were originally collected. Each of the sources of error can potentially contribute both variance and bias to estimates. Another division is into sampling errors and non-sampling error, (the latter also being classified by the different sources of error).

Ideally one attempts to provide measures of errors from each source to arrive at quantifiable estimates of total error. In practice only a few sources of error are routinely measured: response rates, as indicators of non-response bias, and sampling errors are the most common.

1.2 Direct measures of error/quality

Early reporting of the quality of survey data relied heavily on reports of sampling error, partly because it was more easily quantified than many other errors; even for complex sampling designs, computer programs were available which could calculate the sampling errors for a particular data set.

Non-sampling error can also, of course, have considerable impact on the accuracy of estimates, resulting in bias and/or an increase in the variance of estimates. These errors can occur from many sources throughout the process of data collection and processing. Over time, awareness grew of the importance of attempting to measure the impact of non-sampling error on data quality, to allow comparisons to be made of the impact of different sources of non-sampling error.

This information is particularly useful at the commissioning planning/design stage of a survey, where the aim is generally to minimise the impact on data quality of the larger sources of error, having previously identified and measured them. Even if non-sampling errors could be avoided, doing so would be very costly in terms of both expense and delaying the release of the data. Non-sampling errors are now increasingly being regarded not as mistakes, but as the result of conscious decisions to provide accurate, timely data at minimum cost. (Ruddock, 1998, Brackstone, 1996)

Obtaining direct measures of the impact of non-sampling error is not straightforward, however, and special studies are often needed.

Bias

Measuring the bias of an estimate requires knowledge of the true population value and often the estimate itself is the only available source of information. However there are sources which can be used. For example, non-response bias in major household surveys can be estimated by matching survey records at the time of a census of population with census records for responders and non-responders and comparing the characteristics of responding and non-responding households. This has been done in Great Britain (Foster, 1998) and the US (Groves and Couper, 1998). Bias is measured not only to increase awareness for the planning and design of future surveys but also because it can be adjusted for after the data collection phase, for example by reweighting.

Variance

Non-sampling variance can also be measured in special studies , for example by randomly allocating interviewers to households to estimate interviewer variance (O'Muircheartaigh and Campanelli, 1998). This study showed that interviewer variance can be as large as the sampling variance due to the geographical clustering of households in postcode sectors. (The reported sampling variance of an estimate is in reality a combination of the actual sampling variance and interviewer variance).

1.3 Wider definitions of quality

Special studies to obtain direct measures of the effect of non-sampling error are frequently expensive and may be difficult to carry out. In many countries in the last 10 years there have been reductions in research and statistical budgets and a tightening of resources so there is now less likelihood of special studies being set up specifically to measure particular aspects of quality. Such studies are certainly unlikely to be carried out regularly or routinely for most surveys. In order to report on data quality, it is therefore necessary to seek other, less direct, measures. Often non-sampling errors are not measured, but **indicators** of their likely impact on the bias or variance of the estimate are calculated. The most commonly used ones are unit or item response rates. These indicators will indicate that bias exists and give some indication of its likely extent, but they do not measure the magnitude of the bias directly.

Less direct measures of quality are also available by focusing attention on the processes leading to the outputs, rather than the outputs themselves, and on intermediate outputs such as edit failures, as well as final outputs.

Reporting on these wider definitions of quality is discussed in more detail in Section 3.

2. Shift of focus to user perspective

One of the drivers of change in definitions of quality and the reporting of quality over the last 5-10 years is a much greater focus on quality from the customer or user perspective, rather than from the producer's.

One strong element influencing this shift in focus in a number of countries has been increasing competition for the collection and analysis of statistical information. This change has been accompanied by a sharper division of roles between survey customers and suppliers, so that customers have to specify their quality requirements when commissioning survey work and need to be in a position to judge the quality offered when they award a contract.

The way these factors have affected Statistics Sweden was cogently summarised in a recent paper (Andersson, Lindstrom and Lyberg, 1997). They have certainly had a strong effect in Britain. The authors all work in Britain's NSI, the Office for National Statistics (ONS) and four of the five work in ONS's Social Survey Division (SSD), the Government's in-house social survey provider for more than 50 years. For the last 10 years, work on most major Government surveys has been subject to a formal tendering process so that SSD and other survey organisation have had to tender for most of the surveys carried out, in competition with each other. Tenders for survey projects are awarded on the basis of fixed price contracts, so customer requirements, including quality requirements, have to be clearly set out in the tender specification in order to be costed and included in the price.

The change to competitive tendering forced SSD to put much more emphasis on the customers and their needs. To assist with this change, SSD adopted Total Quality Management (TQM) in 1993/94, concentrating on using TQM to change the culture to one of customer focus and continuous improvement. In the rest of ONS, although the competitive element has not been such a strong factor, there has also been a (perhaps more gradual) move towards increased

customer focus with different parts of the organisation using different quality management approaches to assist this move. Many NSIs and survey organisations throughout the world are now working with quality management approaches (Morganstein and Marker, 1997).

This increased emphasis on the customer or user carries several consequences. One is that the user needs to take more responsibility for the quality of data; in the past that responsibility belonged almost entirely to the producers. It may be said that users, either those commissioning the surveys or end users of the data, are not necessarily the best judges of quality but there is no doubt that the user-focused world is here to stay. In that world, quality includes concern for timeliness as well as other traditional aspects of quality. Users are ultimately interested in 'value for money', and want to be able to make trade-offs between the different dimensions of quality. They need to have a clear idea of what the different dimensions are and how they impact on the quality of social research in order to make appropriate judgements of best value for money. This has important consequences for the reporting of quality because it means that producers have a greater responsibility to provide customers/users with the information that enables users to judge the quality of the data. Providing that information is quite a challenge.

3. Reporting quality: product/process measures and other typologies

As discussed in Section 1, most sources of total survey error are difficult to measure routinely, so quality has to be assessed by other means, such as indicators of error or measures taken to reduce the likelihood of error in the course of the survey process. Even subjective assessments of the sources of error and their likely importance, and the procedures taken to minimise them, are useful for those concerned with judging data quality. For example, a description might be provided of the procedures used to develop and test questions to ensure they are understood by respondents and answered in the way intended by the researchers. Another example is item non-response. Levels of item non-response can be quantified and can act as an indicator of non-response error but they can also be interpreted as indicating problems with the question, thus making a more subjective assessment of quality.

So when reporting on survey quality there are sometimes measures or quantifiable information available to report and sometimes all that is available are descriptions of the processes or observations that act as pointers. We suggest that it is helpful to think of the distinction between 'measure or quantify' and 'describe' as more of a continuum than a dichotomy because there are grey areas between quantifying and describing (discussed further in Section 3).

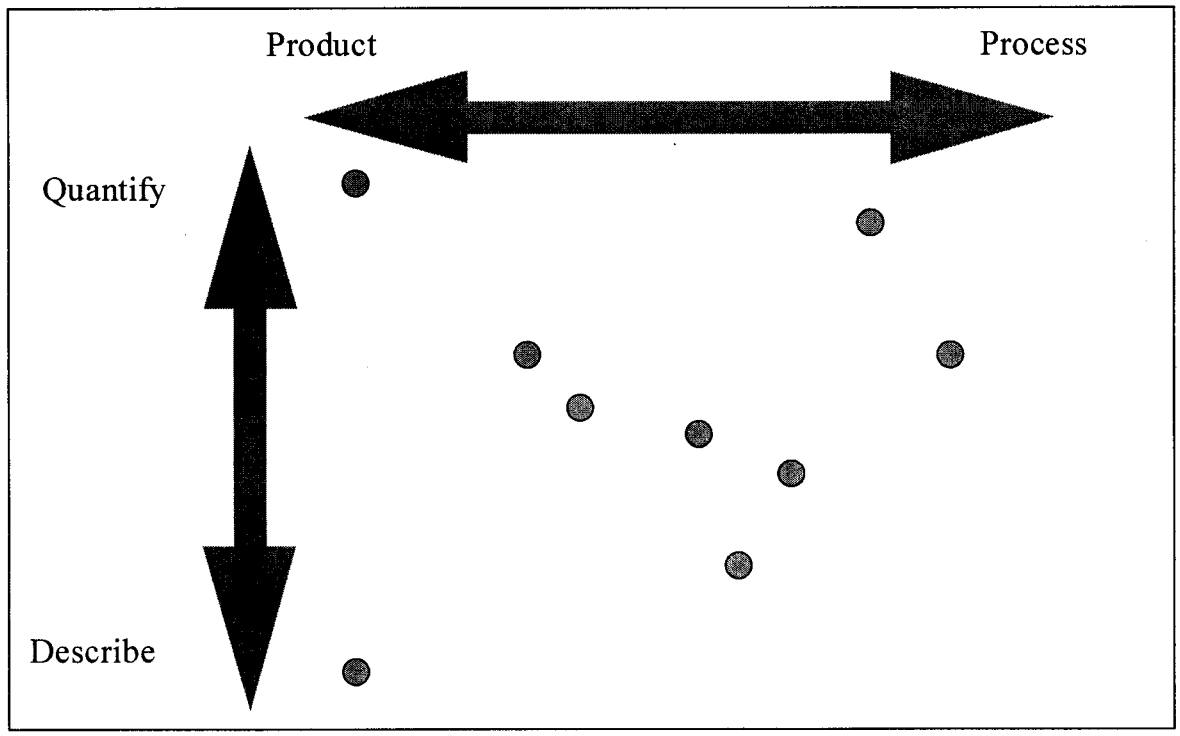
Statistics Sweden (Andersson, Lindstrom and Lyberg, 1997) provide a six-fold classification of the types of information that might be provided when reporting on data quality. The six categories are:

- quantified quality of the product, like evaluation results, variance calculations, response variation.
- quantified process indicators, like non-response and editing rates.
- generalised knowledge on error tendencies from "comparable" surveys.
- process descriptions like coding or editing rules.
- common sense conclusions/vague knowledge - for example about the presence of a black market economy, car accidents not accounted for, etc.
- no knowledge whatsoever of the quality.

The sixth category is a salutary reminder of the fact that survey producers are sometimes right down near the bottom of this classification, with little knowledge about quality - and perhaps should be more willing to report that.

In this classification, quality measures relating to products or outputs are shown as one single category, separate from the second and fourth categories which relate to processes. We find it helpful to show the distinction between product and process quality as another dimension.

The diagram below provides a visual representation of the product/process and quantify/describe dimensions. The dots indicate that quality reporting information may occur anywhere within the matrix. For example, the dot in the top left-hand corner might represent sampling errors, since they quantify the quality of the product. Information about response rates would be towards the top right of the diagram, since they are quantifiable measures but relate primarily to the process rather than the product. The dot part way down the right hand side might represent details of interviewer training – a process measure but one that is partly quantifiable (number of hours/day training, performance in training tests) and partly descriptive (details of the content of the training)



Users are generally interested in the quality of the product, the data. But many of the measures of quality that producers have available relate to the processes leading to the product, rather than the product itself.

For both these dimensions, users often want to be able to compare the quality of one product/survey with another. However, we producers tend not to report on the many aspects of quality in a standard or comparable way. Kasprzyk (1997) noted that survey organisations may develop many different measures of survey data quality for each source of error, and that they communicate their knowledge of error in a variety of ways depending on the nature of the data being released.

So, producers of survey data and estimates are struggling to find better and more standardised ways of reporting on the quality of surveys. Initiatives have included Quality Profiles (Chakrabarty and Torres, 1996; US Energy Information Administration 1996) and the introduction at Statistics Sweden of the Quality Report (Andersson, Lindstrom and Lyberg, 1997). Statistics Canada (1987) has drawn up formal guidelines to specify the quality indicators and assessments that should be provided. Three initiatives on quality reporting being developed within ONS are described in section 4 below. Each classifies or labels the various aspects of quality in a slightly different way.

4. ONS initiatives

4.1 Statistical Quality Checklist

The UK Government Statistical Service (GSS) has recognised the need to provide users with information on the quality of statistics, and about the analytical techniques and processes used to derive the figures, so that they are able to assess the quality of data. This policy is summarised in the GSS Official Statistics Code of Practice:

> "Provide guidance and interpretation to help users understand and use the statistics. This entails informing users of any aspects of the underlying data that could affect the interpretation of the statistics, providing information on accuracy, and making available methods of sampling, collection and analysis for public examination, so that users may make their own interpretations of statistics"

A first attempt to develop this has been the production and publication of a Statistical Quality Checklist (Government Statistical Service, 1997). The publication lists questions which should be considered when presenting statistics in a report or study. Examples of answers for a wide range of surveys are given.

In particular the publication looks at four distinct areas for which the user needs to assess quality:

- The context in which the data was assembled and analysed
- The methods adopted and the limitations these impose
- The reliability of the figures
- The way they relate to other available data on the same subject.

Many of the questions quoted in the Checklist cover process quality and, as such, are not quantified. For example, for data collection methods one of the questions posed is "If the data were collected by interview, what was the training or relevant experience of the interviewers?". The example reply given relates to ONS's General Household Survey and describes the amount of general interviewer training given to all interviewers plus the survey specific briefing given to those working on the survey.

In producing this publication, ONS and the GSS recognise the need to provide the user with considerably more information on data quality, including that which cannot be measured at all or only with considerable effort and expense. Information is provided on the product, and also on the processes involved in data collection and production. As stated earlier, process quality is often a good indicator of data quality when direct measures of the latter are not available.

4.2 Model quality reports for business surveys

The ONS is currently working on a project entitled: **Model quality report in business statistics,** sponsored and funded by Eurostat, as part of their annual research programme[1]. ONS is working in partnership with Statistics Sweden and the Universities of Bath and Southampton in the UK.

A Eurostat working group has been developing models to provide users with the information they need to assess the quality of business statistics. Eurostat define quality in terms of:

- relevance of the statistical concepts;
- accuracy of the estimates;
- timeliness in disseminating results;
- accessibility and clarity of the information;
- comparability of the statistics; and,
- coherence.

This project takes that work forward, considering three of these components: accuracy, comparability and coherence. It aims to demonstrate in practice how quality, especially accuracy, can be assessed in real examples of business statistics, and to develop and implement (with appropriate software where relevant) 'best practice' methodology in the context of specific business surveys undertaken in the two Member States. Four types of output are expected from the project, described in more detail below: a document describing best practice methodology; advice on software, with relevant programs and documentation; guidelines for implementation; model reports for four surveys

[1] The project is due to run for the twelve months of 1998. Final reports are expected to be available by the end of December 1998.

The methodology report

The aim is to provide advice on methods required to assess the accuracy, comparability and coherence of the statistics produced from typical business surveys

The components of accuracy developed by the Eurostat working group use the traditional division of sampling errors and non-sampling errors; the components within each of those headings are shown below

Sampling errors	**Non-sampling errors**
Bias and variance under probability sampling Non-probability sampling	Frame errors Processing errors Non response errors Measurement errors Model assumption errors

Software

Advice will be given on the use of existing variance estimation software[2]. The evaluation criteria being used are:

- suitability for typical business statistics context;
- flexibility and range of design/estimation available;
- ease of use;
- availability/platforms/price; and
- speed and efficiency.

Guidelines for implementation

The implementation report will set out some guidelines on ways in which to implement a quality measurement system. It will be based on the experiences both in Sweden and in the UK of doing this within the project.

Model reports

The final output of the project will consist of four model reports, two based on UK statistics and two on Swedish statistics. The reports will concentrate on assessing accuracy, comparability and coherence of the data, according to the headings set out in the Eurostat documentation. However, the coverage of the reports will also provide users with more information on the statistical products, such as survey objectives, coverage, and estimation techniques.

[2] The following packages will be evaluated: GES (v4.0); CLAN; SUDAAN; WesVarPC; PC-CARP; STATA.

4.3 Developing Standard Quality Indicators for social surveys

SSD carries out a wide range of social surveys for ONS, the UK Government Statistical Service and other public sector bodies. Preliminary work focusing on the reporting of quality led to the realisation that there are surprisingly few tangible measures of quality that are regularly produced in a standard fashion across all surveys carried out by SSD. A project was set up with the broad aims of:

- identifying quality and performance indicators currently produced on surveys
- identifying areas where indicators could be introduced
- recommending a standard list of indicators for use on all surveys and how these indicators can be produced and reported on within SSD in a standard and meaningful way.

The project provoked some discussion about what is a quality indicator and what is a process or progress monitoring tool. There is a considerable overlap. Response rates, for example, are often used as a measure of product quality, though in fact, they are strictly a measure of process which is used as a proxy for the quality of the data - the higher the response rate, the less likely there is to be non-response bias in the survey results. There is no need to be dogmatic about the definitions but it is important that internal monitoring tools are not used as indicators of product quality without proper understanding of what it is that the tool is actually measuring. There is a danger that, because some things are measurable, they are used as quality indicators, without due consideration of whether the measure is meaningful.

The four main components of quality identified by Statistics Sweden (Andersson et al, 1997) were the starting point.

- Content
- Accuracy
- Timeliness
- Accessibility

Examples of how these components are, or could be measured in the SSD context are:

Content - a comparison of survey results with other statistics
Accuracy - levels of non-response at case and individual level, measurement errors etc.
Timeliness - delivery dates of key products
Accessibility - production of suitable documentation, delivery of data to the Essex archive (an external academic archive)

After discussions with the survey project managers, and field staff, recommendations were made for a standard set of quality/progress indicators to be used on surveys.

A number of areas were excluded from the start. It became obvious early on that many of the progress monitoring tools being used are designed for household surveys and are not appropriate for institution, or site-based, surveys. Although it would be possible to suggest indicators for these surveys, they were not included in the project. The long-term aim would also be to produce measures of quality and performance for all stages of a survey. However, the quality of analysis and report writing is difficult to measure, and there were no indicators from any current work, so it has not been included in the current project.

Another area needing consideration is the validation of results against other sources. The proposed indicators (see below) suggest an annual comparison with population estimates, but it will often be relevant to compare key survey results with other sources. Problems can arise here, however, in knowing which source should be used as a comparison and some thought needs to be given to defining 'gold standards' for individual topics (for example, in the UK the Labour Force Survey would fulfill this role in relation to employment statistics).

It was common to find that surveys with specific needs had developed indicators useful to them, but not applicable across all surveys. Examples include the proportions of households where documents are consulted about financial details, and the monitoring of the number and type of diary entries. These survey specific measures do not form part of the list of key measures but will continue to be produced in addition to the standard measures. The standard measures will form a minimum.

The first list of standard indicators produced includes measures which relate to both product and quality; examples are shown in the table below. The frequency with which indicators should be routinely

[3] The following packages will be evaluated: GES (v4.0); CLAN; SUDAAN; WesVarPC; PC-CARP; STATA.

produced needs to be considered; it will vary according to the nature of the survey. For example, for continuous surveys response rates warrant regular, probably monthly, measurement whereas for ad hoc studies one measure at the end of fieldwork is probably sufficient (for quality reporting purposes, more frequent monitoring will be required for field management purposes).

Initial set of SSD standard indicators

Response rates in various detail (monthly)
% households/individuals fully co-operating
% households/individuals partially co-operating (where applicable)
% households/individuals proxy (where applicable)
% households/individuals refusal to headquarters (in response to advance letter sent before interviewer calls)
% households/individuals refusal to interviewer
% households/individuals non-contact
% households/individuals ineligible to survey

% non-response to key questions (quarterly)
project manager to identify key question for individual survey

Interviewer experience (monthly)
Number/percentage of interviewers new to survey
Number/percentage of interviewers new to SSD

Data content
Comparison of age/gender structure of sample with population estimates
Production of frequencies on harmonised questions* for comparisons with other surveys

Accuracy
Complex standard errors of key demographic variables

* ONS, on behalf of the UK GSS, has led a move to standardise questions and variable across different surveys and published an agreed set of harmonised questions (GSS, 1995)

Summary

Producers of survey data and estimates must respond to the increasing demand from users for information that enables different users to assess the quality of the data for their different purposes. Traditional reporting of data quality, based on direct measures of quantifiable error, provides only a partial picture and is not sufficiently helpful to users. A much wider view of quality needs to be taken, with quality reporting including not just quantifiable measures but also descriptive information and not just indicators relating to the outputs but also indicators relating to intermediate outputs and the processes leading to the outputs.

Users may also wish to compare information about the quality of different data sources so it would be desirable to develop more common or standardised methods of quality reporting.

References

Andersson C, Lindstrom H L and Lyberg L (1997) 'Quality Declarations at Statistics Sweden – Principles and Practises'. Statistical Policy Working Paper 26, *Seminar on Statistical Methodology in the Public Service,* Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget

Brackstone G (1996) "Opening remarks" *Proceedings of Statistics Canada Synposium 96, Non-sampling errors,* Ottawa

Chakrabarty R P and Torry G (1996) *American Housing Survey: A Quality Profile*, Current Housing Reports, H21/95-1, Washington DC: US Government Printing Office.

Foster, K (1998) 'Evaluating non-response on household surveys: report of a study linked to the 1991 Census' *GSS Methodology Series*. In press.

Groves R M (1989). *Survey Errors and Survey Costs.* John Wiley.

Groves R M and Couper M P (1998) Non-response in Household Interview Surveys. New York: Wiley.

GSS (1997) 'Statistical Quality Checklist', Office for National Statistics

Lyberg L, Biemer P and Japec, L (1998) 'Quality Improvements in Surveys – A Process Perspective' ASA, Dallas.

Morganstein, D and Marker, D (1997) 'Continuous Quality Improvement in Statistical Agencies' in L.Lyberg, P. Biemer, M. Collins, E. DeLeeuw, C. Dippo, N. Schwarz, and D. Trewin (eds.), *Survey Measurement and Process Quality,* pp 475-500, New York: John Wiley.

O'Murcheartaigh, C and Campanelli, P (1998) 'The relative impact of interviewer effects and sample design effects on survey precision', *Journal of the Royal Statistical Society*, Series A vol 161(I) pp63-77.

Kasprzyk, D and Kalton, G (1997) *Measuring and Reporting the Quality of Survey Data.* Paper presented at Symposium 97, New Directions in Surveys and Censuses, Statistics Canada,Ottawa

Ruddock, Vera (1998) 'Improving and measuring data quality across the Government Statistical Service', *Proceedings of 1998 GSS Methodology Conference*, London

Ruddock, Vera (1998) 'Improving and measuring data quality across the Government Statistical Service', *GSS Methodology Series.* In Press.

Statistics Canada (1987) *Quality Guidelines.* Ottawa, Canada

US Energy Information Administration (1996) *Residential Energy Consumption Quality Profile*, Washington DC: US Department of Energy