ISSN 1977-0375

eurostat
Methodologies and
Working papers

ESS Standard for Quality Reports

2009 edition

eurostat
EUROPEAN COMMISSION

How to obtain EU publications

Our priced publications are available from EU Bookshop (http://bookshop.europa.eu), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents. You can obtain their contact details by sending a fax to (352) 29 29-42758.

Europe Direct is a service to help you find answers to your questions about the European Union

Freephone number (*):

00 800 6 7 8 9 10 11

(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

More information on the European Union is available on the Internet (http://europa.eu).

Luxembourg: Office for Official Publications of the European Communities, 2009

ISBN 978-92-79-07854-5
ISSN 1977-0375
Cat. No. KS-RA-08-015-EN-N

Theme: General and regional statistics
Collection: Metholodogies and working papers

FOREWORD

The role of quality reporting will be strengthened in the forthcoming revision of the basic legal framework for European Statistics, reflecting the importance of quality in the world of statistics. Producers of statistics will have to demonstrate that high quality standards have been applied and achieved through all steps of the statistical production processes. Furthermore, users of statistics will be guaranteed access to appropriate metadata describing the quality of statistical outputs, so that they will be able to interpret and use the statistics correctly.

In 1998 Eurostat set up a Working Group on Quality, comprising members of the European Statistical System (ESS), which has developed and secured agreement on many aspects of the current quality framework, such as the ESS definition of quality and the ESS standard quality reporting documents. In addition, the ESS Leadership Expert Group on Quality, established in 1999, has produced a comprehensive set of recommendations for improvement of the ESS, which have been implemented by National Statistical Institutes through Eurostat-supported projects. This has resulted in the development of process and output oriented quality assessment handbooks and tools, such as the "Development of a Self-Assessment Programme (DESAP)", the "Handbook on Improving Quality by Analysis of Process Variables" and the "Data Quality Assessment Methods and Tools (DatQAM)", all of which are publicly available from the Eurostat Quality website.

The development of the European Statistics Code of Practice was a logical continuation of the focus on quality. It provides a broad conceptual framework for viewing quality and sets standards for the ESS institutional environment, statistical processes and statistical outputs.

The ***ESS Standard for Quality Reports*** will assist National Statistical Institutes and Eurostat in meeting the Code of Practice standards by providing recommendations for preparing comprehensive quality reports for the full range of statistical processes and their outputs. The ***ESS Handbook for Quality Reports*** provides much more detailed guidelines and examples of quality reporting practices. Both documents replace the 2003 versions and contain significant updates reflecting the advances in quality practices over the past five years. In particular, the documents are built around the fifteen principles articulated in the Code of Practice. The coverage of statistical processes using administrative sources or involving multiple data sources has been improved and more quality and performance indicators have been included.

The Standard and the Handbook are applicable to National Statistical Institutes and Eurostat in their roles as producers, compilers and disseminators of statistics. A key objective is to promote harmonised quality reporting across statistical processes and across Member States and hence to facilitate cross-comparisons of processes and outputs.

The new documents have been prepared by two statistical consultants - Michael Colledge and Jörgen Dalén - in cooperation with the members of the Working Group on Quality and staff in Eurostat. I would like to thank all colleagues in the ESS who have contributed to the development and finalisation of the documents.

Walter Radermacher
Director General Eurostat

Table of Contents

Part I: Introduction

The general aim of this Standard (ESQR) is to provide recommendations for preparation of comprehensive quality reports for a full range of statistical processes and their outputs. In this context the term ***statistical process*** means sample survey, census, use of administrative data, production of price or other economic index, or any other statistical compilation commonly performed by a national statistical office, and the term ***national statistical office (NSO)*** refers to the ***national statistical institute (NSI)*** that plays the lead role in a national statistical system or to any other national agency or unit that produces official statistics of relevance to the European Statistical System (ESS).

For more detailed guidelines, including examples, the reader is referred to the ESS Handbook for Quality Reports (EHQR), which also contains definitions of key concepts and references to existing quality reports and other documentation.

A key objective of the Standard is to promote harmonised quality reporting across statistical processes and Member States and hence to facilitate comparisons across processes and outputs. Its focus is on national quality reports and their use in reporting to Eurostat. For guidelines concerning ESS level quality reports, the reader is referred to the EHQR.

The Standard is applicable to NSOs and Eurostat in their roles as producers, compilers and disseminators of statistics. Primarily the Standard is producer-oriented, i.e. aimed at internal assessment of process and output quality. Quality reports based on the Standard will, however, include all the information necessary for user-oriented quality reporting.

For the purpose of the Standard six types of statistical processes are distinguished.

1. ***Sample Survey.*** A survey based on a, usually probabilistic, sampling procedure involving direct collection of data from respondents.
2. ***Census.*** A survey, where all frame units are covered.
3. ***Statistical Process Using Administrative Source(s)***. A process making use of data collected for other purposes than direct production of statistics. An example is where statistical tabulations are produced from an administrative database maintained by the agency responsible for higher education. .
4. ***Statistical Process Involving Multiple Data Sources***. Different sampling, questionnaire designs and/or sampling procedures are used for different survey segments.
5. ***Price or other economic index process***. Involving complex sample surveys, often with non-probabilistic designs, and the target is complex and model-based.
6. ***Statistical Compilation***. Specifically including economic aggregates like the National Accounts and the Balance of Payments.

The standard is organised along the lines of the quality principles in the ESS Code of Practice. Following section (1) introduction, the Standard requires information on (2) relevance, (3) accuracy, (4) timeliness and punctuality, (5) accessibility and clarity, (6) coherence and comparability, (7) output quality trade-offs, (8) user needs and perceptions, (9) performance, cost and respondent burden, (10) confidentiality, transparency and security, and finally (11)

conclusion. The quality and performance indicators recommended for inclusion are listed in the Annex.

A quality report can have narrow or wide scope - from dealing with a specific indicator and the process that produced it to the whole ESS. The Standard refers to the quality aspects of a statistical process (as previously defined) at national level. It is aimed at the most comprehensive form of quality report commonly prepared, i.e., a full scale report with qualitative and quantitative information, dealing in detail with all important aspects of output and process quality. It is envisaged that such a report should exist and to be kept up-to-date, preferably annually, for all statistical processes in the ESS.

As a quality report is a means to an end, not an end in itself, the Standard envisages that a quality report should provide not only a factual account of quality but also recommendations for quality improvements and justification for their implementation.

Part II: Standard

1 Introduction to the Statistical Process and Its Outputs

An introduction is needed to provide context for the quality report.

What should be included in the Introduction

- A brief history of the statistical process and outputs in question.
- The broad statistical domain to which the outputs belong; related statistical outputs.
- The boundary of the quality report at hand and references to related quality reports.
- An overview of all output produced by the statistical process.
- Reference to other documentation, especially on methodology.

2 Relevance

Relevance is the degree to which statistical outputs meet current and potential user needs. It depends on whether all the statistics that are needed are produced and the extent to which concepts used (definitions, classifications etc.,) reflect user needs.

2.1 For All Types of Statistical Process

A ***content-oriented description*** of all statistical output should be given, including key indicators, variables, subdomains, estimates of level and change and reference period.

An assessment regarding the ***key outputs/estimates*** desired by different categories of users should be provided. Shortcomings in outputs for important users should be mentioned. Future plans for improvement should be presented as well as reasons why certain user needs cannot be fully met.

Definitions of ***statistical target concepts*** and their relation to the target definitions that would be ideal from a user perspective should be given. Discrepancies between the definitions used and accepted ESS or international definitions should always be clearly pointed out.

If certain indicators, variables and/or domains foreseen by the ESS or other international regulations/ guidelines are not covered, the statistical outputs are ***incomplete***. An explicit statement of the degree of completeness in terms of ESS regulations should be given where relevant, including plans for future improvements. Completeness can also be measured relative to a national target.

2.2 For Statistical Processes Using Administrative Source(s)

When administrative data are used for statistical purposes, the definitions of the included variables are already fixed based on the primary purpose of the register. These definitions are often not the ideal ones for statistical purposes. The quality report should include definitions of important variables in the register and discuss their relation to the definitions desired by key users and uses of the statistics.

2.3 For Price or Other Economic Index Process

In price indexes the target of estimation is usually not possible to define exactly and is even open to controversy. A quality report should discuss important issues concerning the target of estimation, its relation to approaches and methods chosen and also relate these to recommendations in international manuals or legal documents in the ESS system.

2.4 For Statistical Compilations

The quality report needs to relate the definitions and conceptual choices made to those recommended international manuals or by other forms of general agreement.

What should be included on <u>Relevance</u>

- A content-oriented description of all statistical outputs.
- Definitions of statistical target concepts (population, definition of units and aggregation formula) including discrepancies from ESS/international concepts. (Can also be discussed under Coherence and Comparability.)
- Information on completeness compared with relevant regulations/guidelines.
- Unmet user needs, including reasons for not meeting them.
- Available quality indicators.

3 Accuracy

*The **accuracy** of statistical outputs in the general statistical sense is the degree of closeness of estimates to the true values.*

3.1 For All Types of Statistical Process

A purpose of statistics is to produce estimates of unknown values of quantifiable characteristics of a target population. Estimates are not equal to the true values because of variability and bias.

There are ***sampling errors***, and ***non-sampling errors*** including (i) coverage errors, (ii) measurement errors, (iii) nonresponse errors, and (iv) processing errors.

The section on accuracy in a quality report has to be accompanied by a presentation of the methodology used, which may be included or referenced. The section should normally begin with identification of the main sources of random and systematic error of the statistical output (estimates). Key variables, domains and estimates should be given special focus. The section should include a summary discussion of all errors, random and systematic. For key indicators an assessment of the aggregate risk of random and systematic error should be made.

According to the state of knowledge of the producer, the assessment of bias can be in quantitative or qualitative terms, or both. It should reflect the producer's best current understanding and include actions taken to reduce bias. A qualitative assessment should refer to the likely sign of

the bias and a statement on its order of size, for example by stating its likely maximum value or by using terms like *negligible, small,* or *large*. The information supporting such a statement should be included as well.

Specific sources of error should be described in separate sub-sections under accuracy. Different types of statistical process are affected by different types of errors and the relative importance of each type varies. Therefore the detailed organisation of the section on accuracy in a quality report needs to be unique for each type of process.

What should be included on <u>Overall Accuracy</u>

- A presentation of the methodology sufficient for (i) judging whether it lives up to internationally accepted standards and best practice and (ii) enabling the reader to understand specific error assessments.
- Identification of the main sources of error for the main variables.
- A summary assessment of all sources of error with special focus on the key estimates.
- An assessment of the *potential for bias* (sign and order of magnitude) for each key indicator in quantitative or qualitative terms.

3.2 For Sample Surveys

3.2.1 Sampling errors

For ***probability sampling***, sampling theory provides techniques for the estimation of the expected value and variance of specific indicators over all possible samples, Therefore, the random variation due to sampling can be calculated. Furthermore, sampling biases are normally zero or negligible so that the variance can be taken to represent total sampling error.

Sampling errors should be reported for all estimates resulting from a statistical process where sampling is involved. Where there is a scientific basis for their calculation, they should be given in quantitative terms along with the estimation and variance formulas. There are several presentational devices that can be used, such as standard errors, coefficients of variation (CV) and confidence intervals.

The CV is the most suitable sampling error statistic for quantitative variables with large positive values, which are common in economic statistics. It is not recommended for percentages or changes, where it could easily be misunderstood and it is not usable for estimates that can take on negative values. In these cases a confidence interval is often a better statistic.

For key indicators the sampling error should always be expressed as a confidence interval, since this is the most rigorous and clear way of demonstrating sampling variability.

Especially in economic surveys, ***outliers*** can greatly influence the estimates and lead to major sampling errors. The quality report should state clearly, whether, how and why outlying sample units have received special treatment in the estimation process.

When ***non-probability sampling*** is applied, random error can not be estimated without reference to a model of some kind. Furthermore, sampling biases may well be significant and need to be

assessed as well. There are many types of non-probability sampling, which each require their own evaluation depending on the situation at hand.

For ***cut-off sampling***, frequently applied in economic surveys, the most suitable approach is two-fold. For the sampled portion of the population, random sampling error may be presented as above. For the non-sampled portion a discussion about the (explicit or implicit) model used in the estimation process should be included.

For other forms of non-probability sampling, for example those applied in price indexes, it can be reasonable to invoke a model implying that a sample is "effectively random" according to some design and then apply the relevant variance estimation formula according to that design. It has to be complemented with a discussion of possible sampling bias and of possible limitations in the sampling model used.

It is not enough to just declare that a sample is "purposive" or "subjective" without providing more information. Technical details on how the sample was selected should always be reported or made available. The rationale for not using probability sampling should be stated as well as an assessment of how the sampling procedures can affect the estimates obtained.

Estimates of change are often the most important for users. Although technically more difficult, attempts should therefore be made to also include sampling errors for such estimates. Reasonable assumptions could be used.

What should be included on <u>Sampling Errors</u>

<u>Always applicable</u>

- Where sampling is used there should always be a section on sampling errors. As far as possible sampling error should be presented also for estimates of change in addition to estimates of level. If necessary, reasonable assumptions can be used. If probability sampling is used:
- Presentation of sampling errors, calculated according to formulas, which should also be made available.
- The most appropriate presentational device should be chosen, normally CV, ranges of CV or confidence intervals.
- If outliers have received special treatment in estimation, this must be clearly described.

If probability sampling is used:

- Presentation of sampling errors, calculated according to formulas, which should also be made available.
- The most appropriate presentational device should be chosen, normally CV, ranges of CV or confidence intervals.
- If outliers have received special treatment in estimation, this must be clearly described.

If non-probability sampling is used:

- For cut-off sampling an assessment of sampling bias should be included in addition to the presentation of sampling error for the sampled portion of the population.
- For other forms of non-probability a sampling model could be invoked for the estimation of sampling error. A motivation for the chosen model and a discussion of sampling bias should be included.

3.2.2 Coverage Errors

Coverage errors (or frame errors) are due to divergences between the target population and the frame population. Three types of coverage error are distinguished: (i) undercoverage, (ii) overcoverage and (iii) multiple listings. Another sort of frame deficiency is ***misclassification***, incorrect information about frame units. Such a deficiency causes errors other than coverage errors.

Overcoverage and multiple listings can be handled by statistical methods and will result in an increase of sampling error and cost but no significant biases.

Quantitative information on overcoverage and multiple should be included in the quality report in sufficient detail with respect to, e.g., important sub-domains.

Undercoverage cannot be detected in the measurement process and is the most serious type of coverage error. It will always result in biases, large or small, which are difficult to detect and evaluate. There are methods to detect undercoverage and assess its effects. As far as possible, assessments or estimates of undercoverage should be included in the quality report.

The quality report should include a description of the register or other frame source in so far as this assists in understanding coverage errors and their effects.

What should be included on Coverage Errors

- Quantitative information on overcoverage and multiple listing
- An assessment, preferably quantitative, on the extent of undercoverage and the bias risks associated with it.
- Actions taken for reduction of undercoverage and associated bias risks,
- Information on the frame: reference period, updating actions, references to other documents on frame quality.

3.2.3 Measurement Errors

Measurement errors are errors that occur during data collection and cause recorded values of variables to be different from the true ones. Their causes are commonly categorized as:

- Survey instrument: the form, questionnaire or measuring device used for data collection may lead to the recording of wrong values.
- Respondent: respondents may, consciously or unconsciously, give erroneous data.
- Interviewer: interviewers may influence the answers given by respondents.

Measurement errors may cause both bias and extra variability of statistical outputs. Bias is usually the main problem. The evaluation of measurement errors depends on the type of data.

Data editing identifies inconsistencies in the data, which usually represent errors in the original data. (Such errors could also be processing errors due to coding or data entry.) Information from the data editing process should be included in the quality report, since it gives information about the risk of measurement error. Questionnaires used in the survey should be attached to the quality report as annexes. The efforts made in design and testing the questionnaires should be briefly described.

Important measurement errors are unique for each survey and thus should be presented based on available analyses, or, in the absence of such analyses, the producer's best knowledge.

What should be included on <u>Measurement Errors</u>

- Identification and general assessment of the main risks in terms of measurement error.
- If available, assessments based on comparisons with external data, re-interviews or experiments.
- Information on failure rates during data editing.
- The efforts made in questionnaire design and testing, information on interviewer training and other work on error reduction.
- Questionnaires used should be annexed (if very long by hyperlink)

3.2.4 Nonresponse Errors

Nonresponse is the failure of a sample survey (or a census) to collect data for all data items in the survey questionnaire from all the population units designated for data collection. The difference between the statistics computed from the collected data and those that would be computed if there were no missing values is the **nonresponse error**.

There are two types of nonresponse:

- *unit nonresponse* which occurs when no data are collected about a population unit designated for data collection, and
- *item nonresponse* which occurs when data only on some but not all the survey data items are collected about a designated population unit.

The extent of response (and accordingly of nonresponse) is measured in terms of response rates of two kinds:

- *unit response rate*: the ratio of the number of units for which data for at least some variables have been collected to the total number of units designated for data collection;
- *item response rate*: the ratio of the number of units which have provided data for a given data item to the total number of designated units or to the number of units that have provided data at least for some data items.

Other ratios are sometimes used instead of or as well as these ratios of counts. They are

- *Design-weighted response rates*, which sum the weights of the responding units according to the sample design.
- *Size-weighted response rates*, which sum the values of auxiliary variables multiplied with the design weights, instead of the design weights alone.

The response rates provide indirect indications of the bias risks but the actual bias depends also (and mainly) on the average differences between the respondents and nonrespondents with respect to a survey variable. Normally there is some evidence, although rarely firm, on this matter, which should be included in the quality report in the form of a qualitative assessment.

The increased sampling error can and should be taken into account when computing CVs or confidence intervals.

Efforts and measures, including response modelling, taken to reduce nonresponse in the primary data collection and follow-ups should be described. The technical treatment of nonresponse in the estimation stage (by imputation or by exclusion) should also be clearly stated.

What should be included on <u>Nonresponse Errors</u>

- Nonresponse rates according to the most relevant definitions for the whole survey and for important sub-domains
- Item nonresponse rates for key variables
- A breakdown of nonrespondents according to cause for nonresponse
- A qualitative statement on the bias risks associated with nonresponse
- Measures to reduce nonresponse
- Technical treatment of nonresponse at the estimation stage

3.2.5 Processing Errors

It is necessary to distinguish two very different kinds of processing errors affecting statistical data. The first type concerns *micro-data* where errors could be generated during data entry, data editing (checks and corrections), coding or imputation. For processing error affecting macro-data see Section 4.8.3.

One case where processing error is especially important to evaluate and report is where manual coding of response data in free text format is used. The quality of a coding operation depends in a complex way on the coding rules, how they are interpreted in practice and on the downright mistakes committed by the coders.

Processing errors affecting individual observations cause bias and variation in the resulting statistics, just as measurement errors do. The importance of micro-data processing errors varies greatly between different statistical processes and their treatment in a quality report needs to be proportional to their importance. When they are significant, their extent and impact on the results should be evaluated. If such an evaluation has been made it should be included in the quality report, else there should be a qualitative assessment of the likely impact of processing errors.

What should be included on <u>Processing Errors</u>

- Identification of the main issues regarding processing errors for the statistical process and its outputs.
- Where relevant and available, an analysis of processing errors affecting individual observations should be presented, else a qualitative assessment should be included.

3.3 For Censuses

The objective of a census is to include all units according to an agreed definition. By definition there is no sampling error in a census but non-sampling errors are essentially of the same types.

The following aspects are of special importance for censuses based on extensive field work:

- Risk of undercoverage or overcoverage (usually referred to as undercount and over- or doublecount in the census context). The quality report should assess this source of error, i.e., that field procedures do not reach all target units or that they reach them twice.
- Measurement and nonresponse errors may well be important and the same reporting principles apply to censuses as to sample surveys.
- Processing errors in the form due to the method data entry and coding.

What should be included on <u>Accuracy for a Census</u>

- An evaluation/assessment of undercoverage and overcoverage.
- A description of methods used to correct for undercoverage and overcoverage.
- An evaluation/assessment of measurement and classification errors.
- An evaluation/assessment of processing errors, especially where manual coding of data in free text format is used.

3.4 Statistical Processes Using Administrative Source(s)

For statistics calculated directly from registers, key types of errors are:

(i) *Coverage*. Over- and undercoverage of units that should have been included according to the register definition used should be assessed and reported.

(ii) *Errors in register variables*. For various reasons a register unit may have an erroneous value for a certain variable. This could be compared to a measurement error in a survey.

For event-reporting systems, the quality of data first and foremost depends on the completeness of the reporting system. The *rate of unreported events* (could be regarded as a kind of coverage error) is a key quality factor that needs to be assessed.

What should be included on Accuracy for a Statistical Process Using Administrative Source(s)

- An evaluation/assessment of undercoverage and overcoverage.
- An evaluation/assessment of errors in classification variables.
- For event-reporting systems, an estimate/assessment of the rate of unreported events.

3.5 For Statistical Processes Involving Multiple Data Sources

In many statistical areas, measurement problems are such that the statistical process comprises a number of segments, each of may have its own particular set of units, data sources, questionnaires, sampling scheme, etc. The quality report should focus on the whole picture as well as each of the segments. An early section should contain an overall description of the organisation of the process and its various segments, and a summary of the quality aspects. Then, for each segment, the critical quality/ accuracy aspects should be reported as described in the relevant paragraphs in this Standard.

What should be included on Accuracy for a Statistical Process Involving Multiple Data Sources

- An overall description of the organisation of the survey, the various segments and a summary of the quality aspects.
- For each segment, the items as specified in the appropriate sections in the Standard. (These items may be grouped by segment and/or by error type/quality component.)

3.6 For Price and Other Economic Index Processes

Price indexes are based on statistical surveys and their objective is to monitor price differences in time or space for all products (goods or services) within their scope and to provide an overall estimate of price change/difference. They often involve data from multiple sources and different measurement approaches are often used for different types of products (segments). A quality report needs to describe and assess the approaches for each one of these types of products. Volume and productivity indexes, such as the Industrial Production Index have somewhat similar properties and quality reporting requirements.

Approaches for estimating *sampling error* in price indexes have been pioneered by some countries but no generally agreed approach exists due to the many sampling dimensions (weights, products, outlets/companies). A quality report should include at least a discussion of all relevant sampling dimensions.

Undercoverage likewise has to be reported in all sampling dimensions.

Quality adjustment including replacements and re-sampling needs to be addressed specifically.

Non-response and other errors are normally regarded as secondary problems in price indexes. In a Producer Price Index, it could be more important and should be discussed in a quality report.

What should be included on Accuracy for Price or Other Economic Index Process

- Information on all sampling dimensions (for weights, products, outlets/companies, etc).
- Any attempt at estimating or assessing the sampling error in all or some of these dimensions.
- Quality adjustment methods (including replacement and re-sampling rules) for at least major product groups.
- Assessment of other types of error, where they could have a significant influence.

3.7 For Statistical Compilations

At the top level of the national and European statistical systems there are economic and other aggregates that are compiled from basic statistics from a variety of different sources and that concern various aspects of the economy, society and the environment. The most well known compilations are the National Accounts and the Balance of Payments.

Analysing and reporting the accuracy of these compilations is extremely difficult, since they involve many diverse sources. Special approaches are needed, for example as provided by the IMF's Data Quality Assessment Framework (DQAF) for the National Accounts and the Balance of Payments.

For the National Accounts, *analyses of revisions* are particularly important. A second important topic is the so called *statistical discrepancy*: assessment of the reliability of the consolidated economic and sector accounts can be gained by examination of the capital and financial accounts – which should, in theory, be balanced. Because of errors and omissions in the accounts the balance is rarely achieved. The sizes of the discrepancies are thus indicators of reliability as they reflect errors and omissions.

Analyses of revisions and discrepancies are also appropriate for reporting on the accuracy of the Balance of Payments and other economic compilations.

What should be included on <u>Accuracy for a Statistical Compilation</u>

- Information and indicators relating to accuracy, for example as defined in the IMF's Data Quality Assessment Framework (DQAF) or other relevant, well accepted standard.
- Analysis of revisions between successively published estimates.

<u>For National Accounts</u>

- Analysis of the causes for the statistical discrepancy.
- Assessment of non-observed economy.

3.8 Some Special Issues Concerning Accuracy

3.8.1 Model Assumptions and Associated Errors

There are several types of models in statistical estimation that need to be distinguished.

Model-assisted estimation under a design-based paradigm is used for reducing sampling error, which should be reported. No separate report on model assumptions is needed.

In *model-dependent* estimation there are no design-based estimators to use and the inference depends on the model, whose assumptions need to be critically checked and the basis for the model and its validity analysed and assessed. When used as a remedy for a non-sampling error the discussion of the model should be in the relevant error section rather than in a separate section devoted to the model.

If the *target of estimation is model-based*, the model is usually developed by a subject-matter science. In these cases the model should be described in the quality report and its validity for the data at hand assessed.

What should be included on <u>Model Assumptions and Associated Errors</u>

- Models related to a specific source of error should be presented in the section concerned.
- Domain specific models, for example, as needed to define the target of estimation itself, should be thoroughly described and their validity for the data at hand assessed.

3.8.2 Seasonal adjustment

For statistical processes involving seasonal adjustment, a quality report needs to include a section on this topic. The following points are relevant.

- A short description of the method used, including pre-treatment (calendar effects corrected for, calendar used, type of outliers detected and corrected, model selection and revision and decomposition scheme adopted) and specification of the SA tool chosen (software and version);
- Validation: specification of the quality measures and diagnostics used to evaluate the appropriateness of the identified model and the results of the seasonal adjustment process.

- Revisions: approach chosen for handling revision of seasonally adjusted data in combination or not with revision of raw data (specification of the horizon of revision seasonal factors …).

In case no other documentation is available a full presentation of the process applied and of the methodological choices made, with respect to each item of the ESS guidelines on seasonal adjustment needs to be included in the quality report.

What should be included in the Quality Report on Seasonal Adjustment

- A short description of the method used.
- A report on quality aspects in line with the ESS guidelines on seasonal adjustment.

3.8.3 Imputation

Imputation is a response to deficiencies in the data received. In a sample survey or census the reasons for imputation could be nonresponse (usually item nonresponse) or to correct values affected by measurement or processing errors. The extent to which imputation is used, the reasons for it, and the imputation procedures should be described in the quality report.

What should be included in the Quality Report on Imputation

- Information on the extent to which imputation is used.
- A short description of the methods used and their effects on the estimates.

(Typically this information will be reported in the section(s) dealing with the errors that imputation is helping to correct rather than in a separate section.)

3.8.4 Mistakes

Mistakes are errors in calculation and presentation of the macro-data presented to users. They affect all types of statistical processes in essentially the same way. They are visible to the public and receive a lot of negative attention. Examples are when the methodology is not applied correctly, when the wrong number is inadvertently included in a press release and when analytical presentations or diagrams give wrong impressions. Mistakes can occur in all stages of production of statistics: programming, calculation, the final stage of report writing and editing of manuscripts, etc. The type and number of mistakes that have later been officially revised should be presented for a number of years back.

Procedures applied to minimise risks for gross mistakes in calculation and presentation should be described in the quality report.

What should be included on Mistakes

- The number and nature of mistakes over the past few years should be described.
- Measures taken to avoid mistakes in the future should be described.

3.8.5 Revisions

Planned revisions should follow standard, well-established and transparent procedures. This means for example that pre-announcements are desirable and that the reasons for and nature of

the revision (new source data available, new methods, etc.,) should be made clear. Whether this is the case should be stated in the quality report.

The quality report should first state the relevant revision policy, if there is one, and then present the actual practice. The statement should detail the variables and domains that are subject to revision and the pattern of successive releases.

The quality report should also include information on the size and direction of revisions, and their spread (standard deviation) based on historical data. This information should cover all key indicators.

What should be included on <u>Revisions</u>

- The revision policy.
- The average number of revisions (planned and unplanned).
- The average size of revisions.
- The main reasons for revisions, and the extent to which the revisions improved accuracy.

3.8.6 Subject-dependent techniques on evaluation

For any particular type of statistical process there are unique opportunities for error checking and evaluation. The quality report authors should be creative and take advantage of such opportunities and include the findings in the report. Examples are provided in the EHQR.

4 Timeliness and Punctuality

The ***timeliness*** *of statistical outputs is the length of time between the event or phenomenon they describe and their availability.*

Punctuality *is the time lag between the release date of data and the target date on which they were scheduled for release as announced in an official release calendar, laid down by Regulations or previously agreed among partners.*

Timeliness is relatively easy and straightforward to measure. A common measure is the production time averaged over a number of survey implementations. The maximum production time may also be useful by providing the worst recorded case.

Presentation of punctuality is likewise simple. The most relevant measure is the percentage of releases delivered on time, based on scheduled dissemination dates laid out in Regulations, official timetables or other agreements.

Some statistics are released in several versions, for example preliminary, revised, final. In this case each release has its own timeliness and punctuality profile. The releases should be distinguished and separately presented in the quality report.

Where quality standards have been set up in domain specific regulations and the like they can be used for benchmarking, for example by taking the ratio of, or difference between, the actual production time and the specified production time.

The reasons for possible long production times and non-punctual releases should be explained and the work on improving the situation described.

What should be included on <u>Timeliness and Punctuality</u>

- For annual or more frequent releases: the average production time for each release of data.
- For annual or more frequent releases: the percentage of releases delivered on time, based on scheduled release dates.
- The reasons for non-punctual releases explained

5 Accessibility and Clarity

Accessibility and clarity refer to the simplicity and ease with which users can access statistics, with the appropriate supporting information and assistance.

Evaluation of accessibility can take many forms since accessibility is affected by the many aspects of dissemination practice, including the dissemination channels, the form of the available datasets (microdata or aggregates figures), and pricing policies. The quality report should include a description of the various ways the statistical output can be accessed - by paper, Internet, etc. Pricing policies should be described, together with possible restrictions due to confidentiality or other reasons.

Clarity depends upon the quality of statistical metadata, which are disseminated alongside the statistical outputs. A summary description of these metadata (documentation, explanation, quality limitations, etc.,) should be included in the quality report.

A division of users into occasional users and more experienced, professional users is helpful especially for web-based publishing. Occasional users typically prefer data in static formats that are easy to find and interpret. Experienced users prefer databases, where they can select and download data of interest to them, sometimes for further data manipulation and analysis, together with the corresponding metadata. The quality report should normally refer to the needs of all these kinds of users and how well they have been addressed.

Recent and planned improvements to accessibility and clarity should also be described.

What should be included on <u>Accessibility and Clarity</u>

- A description of the conditions of access to data: media, support, pricing policies, possible restrictions etc.
- A summary description of the information (metadata) accompanying the statistics (documentation, explanation, quality limitations, etc).
- The description should refer to both less sophisticated and more advanced users and how their needs have been taken into account.
- A summary of user feedback on accessibility and clarity.

6 Coherence and Comparability

*The **coherence** of two or more statistical outputs refers to the degree to which the statistical processes by which they were generated used the same concepts - classifications, definitions, and target populations – and harmonised methods. Coherent statistical outputs have the potential to be validly combined and used jointly. Examples of joint use are where the statistical outputs refer to the same population, reference period and region but comprise different sets of data items (say, employment data and production data) or where they comprise the same data items (say, employment data) but for different reference periods, regions, or other domains. **Comparability** is a special case of coherence and refers to the latter example above where the statistical outputs refer to the same data items and the aim of combining them is to make comparisons over time, or across regions, or across other domains.*

The lack of coherence/ comparability of the outputs of statistical processes may thus be summarised under two broad headings - differences in ***concepts***, and differences in ***methods***. The cause of any lack of coherence/comparability, whether due to changes in concepts or methods or both, should be clearly explained. In this situation the producer should facilitate reconciliation of the estimates by quantifying, at least approximately, the effects of the main sources of incoherence/ incomparability. The minimum requirement is that each instance is indicated in the quality report and that the reason for it and its order of magnitude is stated according to the producer's best knowledge.

When bringing together estimates from two statistical processes or the same process over time or across regions, the numerical consistency of estimates depends upon two factors: the coherence/ comparability of the processes; and the errors that actually occurred in generating the estimates, i.e., the accuracy of the processes. Thus, coherence/comparability is a prerequisite for numerical consistency. It determines the potential for numerical consistency; it does not guarantee numerical consistency as this also depends upon errors, which the quality report should report in the section on accuracy.

There are several types of coherence/comparability that are worth distinguishing and which are summarised in the box below. The EHQR provides more details.

What should be included on <u>Coherence and Comparability</u>

<u>General</u>

- Brief descriptions of all conceptual and methodological metadata elements that could affect coherence/ comparability.
- An assessment (preferably quantitative) of the possible effect of each reported difference on the output values.
- Differences between the statistical process and the corresponding European regulations/ standards and/or international standards (if any).

<u>Comparability over Time</u>

- Reference periods when series breaks (if any) occurred, the reasons for them and treatments of them.

Comparability over Region

- A quantitative assessment of comparability across regions based on the (weighted) number of differences in metadata elements.
- At ESS level, a coherence/comparability matrix summarising the possible sources of lack of comparability with any specified standard by region.

Internal Coherence

- Any lack of coherence in the output of the statistical process itself.

Coherence with National Accounts

- Where relevant, the results of comparisons with National Account framework and feedback from National Accounts with respect to coherence and accuracy problems.

Coherence with Other Statistics

- Where the statistical outputs were combined with those from other processes, the limitations if any set by coherence considerations

Mirror Statistics

- Assessment of discrepancies (if any).

7 Trade-offs between Output Quality Components

As previously noted, the output quality components are not mutually exclusive in the sense that there are relationships between the factors that contribute to them. There are cases where the factors contributing to improvements with respect to one component lead to deterioration with respect to another. The report should deal with the trade-offs that have to be made assuming the budget is fixed and that all possible efforts have been made to improve quality with respect to each component individually. The following paragraphs indicate the types of trade-offs that are most likely to require consideration, with examples.

- *Trade-off between Relevance and Accuracy.* For example, relevance may improved by increasing the number of data items collected but time available for editing each data item will be reduced and the potential for measurement errors increased, thus adversely affecting accuracy.
- *Trade-off between Relevance and Timeliness.* Likewise, timeliness can be improved by reducing the number of data items being collected and processed and/or by replacing some of the items that are more difficult to report or process by ones that are easier. This will have a negative effect on relevance.
- *Trade-off between Relevance and Coherence.* Improvements in relevance of a particular statistical process in response to user requirements, for example by fine-tuning the definitions of some variables or classifications may reduce the coherence of its outputs with those of other processes.

- *Trade-off between Relevance and Comparability over Time.* Improvements in relevance in response to user requirements, for example by redefining the items for which data are collected, or moving to a later version of a classification, will reduce comparability over time, perhaps to the point of requiring a series break.
- *Trade-off between Comparability over Region and Comparability across Time.* In a similar fashion, the desire to have more comparability across region may well result in changes that reduce comparability over time.
- *Trade-off between Accuracy and Timeliness.* This is probably the most frequently occurring and important of the trade-offs. Improvement in timeliness can be obtained by reducing collection and processing time, in particular by terminating collection earlier, compiling outputs based on a smaller number of responses and/or by reducing the amount of editing. However, this reduces accuracy.

What should be included on <u>Trade-offs</u>

- A description of each important trade-off that has been analysed and the basis on which the trade-off decision has been made.
- A statement concerning any trade-offs that should have been analysed but have not been.

8 Assessment of User Needs and Perceptions

We provide our users with products and services that meet their needs. The articulated and non-articulated needs, demands and expectations of external and internal users will guide the ESS, its members, their employees and operations (*ESS Quality Declaration - User Focus).*

Assessment of user needs generally involves a classification of users, an indication of the uses for which they want the outputs, the priorities in satisfying their needs, and an account of how this information was obtained, for example through advisory committees, other regularly convened user groups, ad hoc focus groups, feedback/complaints from users.

The most effective method of evaluating user perceptions is a full scale user satisfaction survey. Other, cheaper methods of assessment include analysis of publication sales, user comments, requests and complaints received, web site accesses, etc., and feedback from advisory committees and focus groups.

What should be included on User Needs and Perceptions

- Means of obtaining information on users and uses.
- Description and classification of users.
- Uses for which users want the outputs.
- Users and uses given special consideration.
- Means of obtaining user views.
- Main results regarding user satisfaction.
- Date of most recent user satisfaction survey.

9 Performance, Cost and Respondent Burden

Resources must be effectively used. Respondent burden should be proportional to the needs of users and not excessive for respondents. Respondent burden should be measured and targets set for its reduction over time (European code of Practice, Principles 9 and 10).

The capacity to calculate costs is essential for quality and performance assessment. Cost benefit analyses are required in order to determine the appropriate trade-off between costs on the one hand and benefits in terms of output data quality on the other. (Note that, in this context, sources of funding are irrelevant as they have nothing to do with efficiency.)

Likewise, respondent participation must be viewed as a cost (to respondents) that has to be balanced against the benefits of the data thus provided.

A comprehensive assessment of all the costs associated with a statistical process is complicated because it requires a mechanism for allocating shared costs (for example, the costs of the business register) and overheads (office space, utility bills etc). A simple assessment of the principal direct costs - for example staff, data collection and processing costs - is also feasible.

The response burden carried by respondents, whether individuals, household members or businesses can be measured in hours spent or in financial terms. A commonly used method for business surveys is to measure respondent burden as the product of the number of dispatched or completed questionnaires multiplied by the average completion time. The latter may be estimated using focus groups or based on the responses to a specific question at the end of each questionnaire asking about completion time. Respondent burden in financial terms may be obtained by multiplying by the average hourly cost of respondent's time, again estimated using focus groups.

What should be included on <u>Performance, Cost and Respondent Burden</u>

<u>Performance and Cost</u>

- Annual operational cost with breakdown by major cost component.
- Recent efforts made to improve efficiency.
- The procedures for internal assessment and for independent external assessment of efficiency.
- The extent to which routine operations, in particular data capture, coding, validation and imputation, are automated.
- The extent to which ICT is effectively used for used for data collection and dissemination and the improvements that could be made.

<u>Respondent Burden</u>

- Annual respondent burden in financial terms and/or hours.
- Respondent burden reduction targets.
- Recent efforts made to reduce respondent burden.
- Whether the range and detail of data collected by survey is limited to what is absolutely necessary.
- Whether administrative and other survey sources are used to the fullest extent possible.
- The extent to which data sought from businesses is readily available from their accounts.
- Whether electronic means are used to facilitate data collection.
- Whether best estimates and approximations are accepted when exact details are not readily available.
- Whether reporting burden on individual respondents is limited to the extent possible by minimizing the overlap with other surveys.

10 Confidentiality, Transparency and Security

The privacy of data providers (households, enterprises, administrations and other respondents), the confidentiality of the information they provide and its use only for statistical purposes must be absolutely guaranteed. Statistical authorities must produce and disseminate European statistics respecting scientific independence and in an objective, professional and transparent manner in which all users are treated equitably. (European code of Practice, Principles 5 and 6).

Typically, confidentiality is required by law and survey staff have legal confidentiality commitments. The quality report should confirm such arrangements, or report on any exceptions. It should also outline the procedures for ensuring confidentiality during collection, processing and dissemination. These include protocols for ensuring that individual data are accessed strictly on a *need to know* basis, rules for defining confidential cells in output tables, and procedures for detecting and preventing residual disclosure. In addition, the arrangements, if any, under which users outside the NSO may access microdata for research purposes, and the confidentiality provisions, should be described.

The quality report should also indicate if and how respondents are informed of the mandate under which a survey is being conducted, the uses to which the data requested are to be put, and the confidentiality provisions. It should draw attention to any deficiencies. It should comment on whether statistical announcements and statements made in press conferences are objective and non-partisan, and errors discovered in published statistics are immediately corrected and publicised. It should describe the provisions in place to ensure the security of data acquisition processes, in particular data collected through the Internet, and of the databases in which data are stored. It should also review the trade-offs between confidentiality and security on the one hand and accessibility on the other.

What should be included on Confidentiality, Transparency and Security

Confidentiality

- Whether or not confidentiality is required by law and if so whether survey staff have signed legal confidentiality commitments.
- Whether external users may access micro-data for research purposes, and, if so, the confidentiality provisions that are applied.
- The procedures for ensuring confidentiality during collection, processing and dissemination, including rules for determining confidential cells in output tables and procedures for detecting and preventing residual disclosure.

Transparency

- The ways in which the uses to which the data are to be put and confidentiality provisions are made known to respondents.
- Whether statistical announcements and statements made in press conferences are objective and non-partisan.
- Whether errors discovered in published statistics are corrected and publicised.

Security

- The provisions in place to ensure the security of data acquisition processes, in particular data collected through the Internet.
- The provisions that are in place to ensure the security and integrity of completed questionnaires, micro and macro databases and data outputs.

11 Conclusion

The quality report should conclude with a section referring to the principal quality problems, each of which should be accompanied by recommendations for improvements. The intended users and uses of the quality report and the follow-up action items should be indicated.

What should be included in the Conclusion

- Principal quality problems.
- Recommendations for improvements.
- Follow-up action items.

Annex: Quality and Performance Indicators

The quality and performance indicators recommended for inclusion in quality reports are listed in the table below. More details are provided in the EHQR.

Identification and Name	Brief Description
Relevance	
R1. Rate of available statistics.	The ratio of the number of output data elements provided in accordance with a relevant ESS regulation to those required by the regulation.
Accuracy	
A1. Coefficient of variation (CV).	The standard error of the estimator divided by the expected value of the estimator.
A2. Rate of overcoverage.	The proportion of units accessible via the frame that do not belong to the target population.
A3. Edit failure rate.	The proportion of responding units for which an error signal is triggered by a specified checking algorithm.
A4. Unit response rate.	The ratio of the number of units for which data for at least some variables have been collected to the total number of units designated for data collection.
A5. Item response rate.	The ratio of the number of units which have provided data for a given variable to the total number of designated units.
A6. Imputation rate.	The ratio of the number of assigned values (data are missing, invalid or inconsistent or have failed edits) for a given variable to the total number of values.
A7. Number of mistakes made, by type.	The number of serious mistakes in calculation or presentation of aggregates that are not found until after publication.
A8. Average size of revisions.	The average over a time period of the difference between a later and an earlier estimate expressed as the average revision, the average absolute revision, and/or the corresponding relative quantity(ies).

Identification and Name	Brief Description
Timeliness and Punctuality	
T1. Time lag between end of reference period and date of first/ provisional results.	The number of days from the last day of the reference period to the day of publication of first results.
T2. Time lag between the end of reference period and date of final results.	The number of days from the last day of the reference period to the day of publication of final results.
T3. Punctuality of publication.	The number of days separating a previously announced date of publication and the actual date.
Accessibility and Clarity	
AC1. Number of subscriptions/ purchases of each of the key paper reports.	As stated.
AC2. Number of accesses to on-line databases.	As stated (to be further defined in collaboration with an IT expert).
AC3. Rate of completeness of metadata	The ratio of the number of metadata elements provided to the total number metadata elements applicable.
Coherence and Comparability	
CC1. Lengths of comparable time series.	Number of reference periods in time series from last break.
CC2. Asymmetries for statistics mirror flows.	Discrepancies between data related to flows, e.g. for pairs of countries.
Assessment of User Needs and Perceptions	
US1. User satisfaction index.	The degree of satisfaction with services and products for different segments of users.
US2. Length of time since most recent user satisfaction survey.	As stated.
Performance Cost and Respondent Burden	
PCR1. Annual operational cost, with breakdown by major cost components.	Direct costs of staff involved in data collection (questionnaires, distribution, capture), reducing non-response, processing, and compilation of estimates.
PCR2. Annual respondent burden in hours and/or financial terms	Respondent burden in hours is defined as number of respondents/questionnaires * average time per respondent, summed over all

Identification and Name	Brief Description
	repetitions of statistical process within a year. Respondent burden in financial terms is defined as respondent burden in hours * average hourly cost to respondents.

European Commission

ESS Standard for Quality Reports

Luxembourg: Office for Official Publications of the European Communities

2009 — 28 pp. — 21 x 29.7 cm

ISBN 978-92-79-07854-5
ISSN 1977-0375

KS-RA-08-015-EN-N

Publications Office
Publications.europa.eu