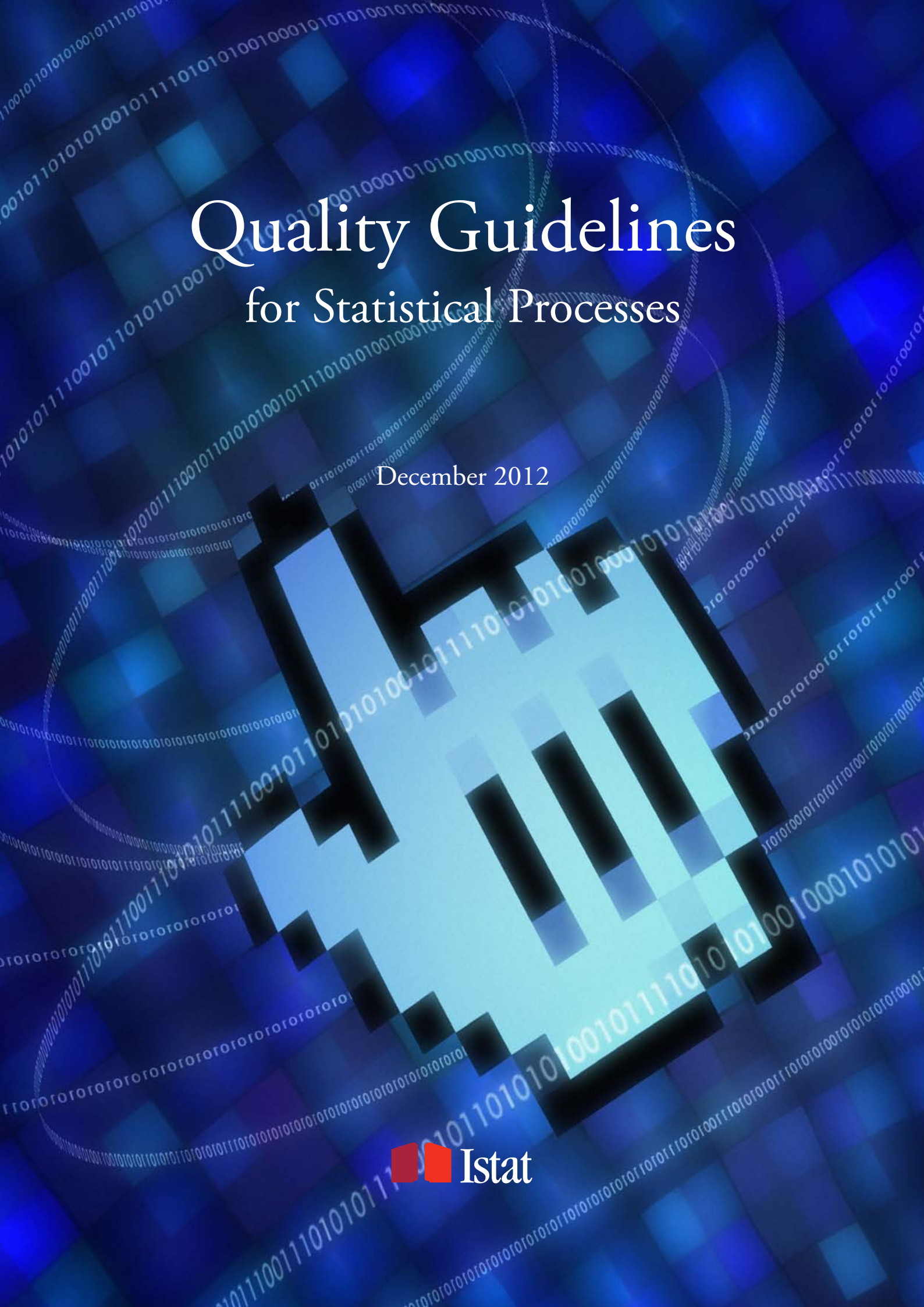
Quality Guidelines
for Statistical Processes

December 2012



Istat

Authors
Marina Signore, Giovanna Brancato, Riccardo Carbini, Marcello D'Orazio, Giorgia Simeoni.

Acknowledgements
The authors would like to thank the following colleagues for their contributions:
Istat Quality Committee (Vittoria Buratta, Fabio Crescenzi, Francesca Di Palma, Stefano Falorsi, Marina Gandolfo, Giuseppe Garofalo, Orietta Luzi, Raffaele Malizia, Silvia Montagna, Alessandro Pallara, Susanna Peddes, Fabio Rapiti, Maria Clelia Romano, Cataldo Scarnera, Susanna Terracina), Fabio Bacchini, Giulio Barcaroli, Mara Cammarrota, Anna Ciammola, Marco Di Zio, Ugo Guarnera, Roberto Iannaccone, Stefania Macchia, Antonia Manzari, Manuela Murgia.

Quality Guidelines
for Statistical Processes

Version 1.1

December 2012

Italian National Institute of Statistics - Istat
Rome

Editing of the English version
Laura Peci

ISBN 978-88-458-1742-7

Index

Introduction

Istat has launched a systematic approach to quality following International and European standards ever since the '90s.

Istat reference framework for quality policies relies on the *European Statistics Code of Practice*, adopted in 2005 and revised in 2011, on Eurostat *Quality Definition* and on the recommendations of the *LEG on Quality*, approved by the Members States of the European Union in 2001. The *Data Quality Assessment Framework*, developed by the International Monetary Fund, also represents an important reference, especially for economic statistics and for National Accounts.

Following the principles of the *European Statistics Code of Practice*, Italy has recently adopted the Italian Code of Official Statistics, O.G. n. 240 of 13/10/2010, in order to promote quality improvements of the statistics produced by the Italian National Statistical System.

In accordance with Eurostat Quality Definition, the requirements of statistical products are:

- to be relevant with regard to users' information needs;
- to be accurate, that is to provide estimates or indicators that are as reliable as possible;
- to timely catch the phenomena to be observed;
- to be easily accessible and supported by metadata allowing for a fully understanding of data;
- to enable comparisons over time or among different sources.

The **Quality Guidelines for Statistical Processes** aim at describing the principles to be followed when planning, running and assessing a statistical process, as well as at illustrating quality requirements of statistics. The *Quality Guidelines* are composed of two parts. The first part is dedicated to Process Quality and follows the phases of the statistical production process. For each phase, the principle or target to be achieved is stated and it is accompanied by summary instructions or guidelines to be followed in order to accomplish it. The second part concerns Product Quality. It describes and explains Eurostat quality requirements, adopted by Istat in order to measure quality and to communicate it to users. However, it does not contain guidelines for measuring quality. Summary information on error sources and their impact on survey results can be found in the first part, while references are provided for more detailed descriptions. Indeed, measuring Product Quality requires to carry out control surveys or experimental studies suitably designed for the specific error source and error type to be assessed.

The *Quality Guidelines* are addressed to survey managers responsible for statistical production and contain the principles identified by Istat in order to guarantee the quality of the statistics produced and disseminated. They constitute the benchmark for assessing process and product quality as well as the degree of compliance with European and national standards by means of internal statistical audit and self-assessment. In both cases, the quality assessment is based on ascertaining the degree of compliance of statistical processes and products with the principles and requirements stated in the *Quality Guidelines*. The quality assessment concerns not only the execution of statistical processes but also quality measurement, documentation and communication to users. For these reasons, the *Quality Guidelines* represent a key reference also for auditors and reviewers[1] in charge of assessing quality.

It has been deemed valuable to issue summary - although comprehensive - *Quality Guidelines* which are intended to be easy and fast to read. The reader looking for detailed descriptions and thorough explanations is addressed to scientific literature and in particular to the *European Manuals of Recommended Practices*.

[1] Reviewers are internal experts in charge of checking the output of self-assessment.

The *Quality Guidelines* reflect the experience and know-how so far achieved by Istat, as well as by other National Statistical Institutes and International Organisations.

Section I: Process Quality

A. Survey Objectives

A.1. Information needs, users and uses

Principle A.1. Information needs, users and uses
The information needs to be met should be clearly defined. At the same time potential users and uses of survey results should be identified. Therefore, surveys should be designed in order to meet the needs of main users.

Guidelines

The survey objectives should be clearly specified and written down during the planning (or re-planning) phase of the survey. More specifically, the following items should be explicitly described in a document: survey objectives, information needs to be met, specific requirements such as content, concepts, periodicity and quality targets, and expected uses of survey results. It is also recommended to clearly point out the legislative acts from which the statistical process originates (e.g. EU regulation or directive, national law). Finally, the need to carry out the survey should be adequately justified with regard to both costs and response burden.

The information needs to be satisfied are determined by current and potential users of survey results. Therefore, main users should be clearly identified and involved in defining objectives and in survey planning (or re-planning).

In general, users are heterogeneous and they often have conflicting interests. For this reason, it is important not only to know the various user types, but also to be able to rank user relevance with respect to survey results. It is useful to produce and regularly update documentation on main users and their characteristics.

Consultation of main users, in order to identify current and potential information needs, may be carried out in various ways, with various levels of formalisation and involvement.

Regular meetings between users and producers called *Quality Circles* took place in the past for planning statistical activities of the Italian National Statistical System. Recently, the Institute, in accordance with decisions taken by COMSTAT[2], set up the National Committee of Users of Statistical Information (CNUIS), in order to enhance the involvement of users during collection, processing and harmonisation of statistical information demand, as well as to define the priorities of the National Statistical Programme (PSN).

Other tools for consultation, such as exploratory surveys on large groups of users, or focus groups on limited number of users, may be arranged *ad hoc* as needed. Finally, research protocols and agreements are examples of high involvement of privileged users who may contribute to the survey both thematically and financially.

In addition to the above mentioned tools for consultation, other channels of contact with main users should be activated and maintained, allowing also to identify information needs not yet satisfied and to

[2] COMSTAT (Policy-making and Co-ordinating Committee for Statistical information) is the governing body of the National Statistical System; it performs steering functions in relation to statistical offices and resolves upon the National Statistical Programme.

anticipate future ones. For example, participation of Istat staff to thematic conferences promoted by users or organization of bilateral meetings or workshops with user groups should be fostered.

In order to identify information needs, indirect methods, based on the analysis of the already available information, should also be used. For example, user requests which was not yet possible to satisfy may be analyzed.

At the end of the survey process, direct tools, such as user satisfaction surveys, or indirect tools, for example the analysis of performance indicators (e.g. number of publications required, number of downloads) should be used in order to evaluate the level of user satisfaction with respect to the produced results and their relevance.

Referring to continuous quality improvement, the main unmet user needs should always be borne in mind and the implementation of projects aiming at satisfying them should be promoted. Survey objectives, contents and procedures should therefore be regularly reviewed, while it is necessary to tend to improving information supply.

Some references

OMB (2006) Standards and Guidelines for Statistical Surveys. Office for Management and Budget, The White House, Washington, USA. http://www.whitehouse.gov/sites/default/files/omb/assets/omb/inforeg/statpolicy/standards_stat_surveys.pdf

Statistics Canada (2009) *Statistics Canada Quality Guidelines Fifth Edition – October 2009*, Catalogue no. 12-539-X, Ottawa. http://www.statcan.gc.ca/pub/12-539-x/12-539-x2009001-eng.pdf

B. Survey design

B.1. Response burden

Principle B.1. Response burden
Particular attention should be paid to minimising response burden. If possible, tools to evaluate response burden should be developed.

Guidelines

Survey planning or re-planning should be finalised to ensure that both the overall survey design and the tools used, in particular the questionnaire, are such as to minimise response burden, while ensuring high quality statistics.

For this purpose, different methods may be applied, according to survey contents, survey objective, survey design and reporting unit. For instance, sampling coordination techniques may be used or questionnaires with pre-printed answers may be provided.

Some techniques, commonly used to promote participation in the survey and reduce non-responses, may also be useful in order to reduce response burden, as, for example, the support provided by well-trained interviewers or the availability of a toll-free number.

An important element for the reduction of response burden is to eliminate questions from the questionnaire which do not contribute to data dissemination. The inclusion of questions processed only for purposes of quality assurance should be carefully evaluated and, in any case, limited.

The measurement or at least the estimation of response burden is essential for planning its reduction. The measurement of statistical response burden is a part of the wider context of the measurement of administrative burden on businesses. Simple indicators of response burden are primarily based on the time required to fill in the questionnaire, and, secondly, on the evaluation of the difficulty faced in finding the required information. This assessment does not consider response burden associated with the request for information on sensitive topics, which is more closely related to the difficulty of dealing with such topics by the respondent, than to the length or complexity of the questionnaire or the difficulty of the information retrieval.

The required time to fill in the questionnaire may be evaluated internally in the production unit, otherwise during the testing phase of the questionnaire (questionnaire test) or of the survey procedures (pilot survey). The required time may also be measured during the data collection phase, for example by asking respondents or the interviewer to specify how much time was required; otherwise it may be automatically registered if computer assisted data collection techniques are used.

An excessive duration of the interview (or of the length of the questionnaire) may cause respondent dropout or the provision of approximate and hasty responses; a long interview should be motivated by specific information needs.

Finally, the level of difficulty for respondents in gathering the required information, in particular for surveys on businesses and institutions, should be assessed. The level of difficulty is low if information

can be easily provided, such as, in most cases, for household surveys; the level is moderate if some kind of information has to be retrieved in documents or in archives; the level is high if a complex search or treatment is necessary to gather the required information, that is, for instance, the case of data referred to past years (that are available in archives), or if the required information is defined using different definitions/concepts with respect to those currently used by respondents.

Some references

Hedlin D., Dale T., Haraldsen G., Jones J. (2005) Developing Methods for Assessing Perceived Response Burden. Report of a project funded by Eurostat http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/DEVELOPING%20METHODS%20FOR%20ASSESSING%20PERCEIVED%20RESPONSE%20BURD.pdf

Dale T., Haraldsen G. (Eds.) (2007) Handbook for Monitoring and Evaluating Business Survey Response Burdens. Eurostat http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/HANDBOOK%20FOR%20MONITORING%20AND%20EVALUATING%20BUSINESS%20SURVEY%20R.pdf

B.2. Survey design for total surveys

Principle B.2. Survey design for total surveys
The decision to conduct a total survey should be justified. The impact of this choice, with respect to a sample survey, should be evaluated in terms of costs, response burden, accuracy and level of detail of the final estimates.

Guidelines

The decision to conduct a total survey rather than a sample survey may be determined by several reasons. For instance, if the target population is relatively small, a total survey may be an adequate choice; the same, if estimates for small study domains are required.

Observing all units obviously determines higher response burden, especially when the same population is involved in other surveys during the same period.

The results of a total survey may be less accurate than those of a well-done sample survey, especially in case of large-scale total surveys where the incidence of non-sampling errors may be very high. Consequently, the survey should be designed so that the impact of non-sampling errors is minimized and the impact of main errors, which has not been possible to avoid, might be evaluated *ex post*.

B.3. Sample design

Principle B.3. Sample design
Sample selection should be carried out according to well-defined probabilistic criteria. The use of non-probabilistic methods for sample selection should be justified. Sample design and sample size should be such so as to ensure a predefined accuracy level for the key variables in the main domains.

Guidelines

Non-probability sampling
The choice of a non-probability sampling design should be justified both theoretically and empirically.

Inferences on a the target population based on a non-probability sample may be misleading, since the risk that estimates may be biased (selection bias) is high. In these circumstances, it is therefore necessary to clarify the assumptions justifying the sample representativeness and to calculate both the estimates of population parameters and the related sampling error estimates.

All aspects of sample design should be well documented.

Probability sampling
The sample design should be adequately defined with respect to survey objectives; it should be as simple as possible; it should ensure that each unit of the sampling frame has a non-zero inclusion probability in the sample (in the case of multi-stage designs, this should be valid for each stage).

The sampling frame should be clearly defined, evaluating its consistency with the objectives of the survey (see also Section C.1).

The sampling design should provide a stratification of the units to create strata of homogeneous units with respect to the information to be collected and to allow the main domains to be obtained by the union of elementary strata.

The optimal sample size should be determined by statistical methods to ensure an adequate accuracy of the estimates for the main variables at the level and for the main study domains. If a substantial reduction of the sample size is expected due to a high number of ineligible units or unit nonresponse, it may be useful to increase the number of the selected sample units.

The sampling design should allow for the estimation of sampling errors (sampling variance).

The task of sample selection from the sampling frame should be carried out by using a generalized software[3]. The use of *ad hoc* software should be limited to peculiar cases, and it should be preliminarily and accurately tested, to avoid that programming errors could invalidate sample randomness.

In general, it is advisable to analyse alternative sample designs, to evaluate pros and cons of each one and to document the reasons of the choice of a specific design.

[3] Details about generalized software used in Istat are available in Istat website, in the section ”Methods and software” http://www.istat.it/it/strumenti/metodi-e-software/software

All aspects of sampling design should be properly documented.

Some references

Särndal C.E., Swensson B., Wretman J. (1992) *Model Assisted Survey Sampling*. Springer-Verlag, New York.

Statistics Canada (2010) *Survey Methods and Practices.* Statistics Canada, Catalogue no. 12-587-X, Ottawa.
http://www.statcan.gc.ca/pub/12-587-x/12-587-x2003001-eng.htm

B.4. Questionnaire design and testing

Principle B.4.1. Questionnaire design
Questionnaires should be designed in order to: effectively collect the information of interest, contain only the necessary questions, do not result in excessive respondent burden and be structured to ensure interview fluency. Whenever possible, standard definitions and classifications - or systems that may be mapped to them - should be used. Questionnaires should be designed so as to facilitate data processing, for example data entry and coding. The graphical layout of paper questionnaires should convey positive perceptions and help user orienting among the different sections. The graphics chosen should be used consistently within the questionnaire.

Principle B.4.2. Electronic questionnaire design and testing
Electronic questionnaires should be developed exploiting the technological potential for routing management and on-line quality control, without unduly burdening the course of the interview.

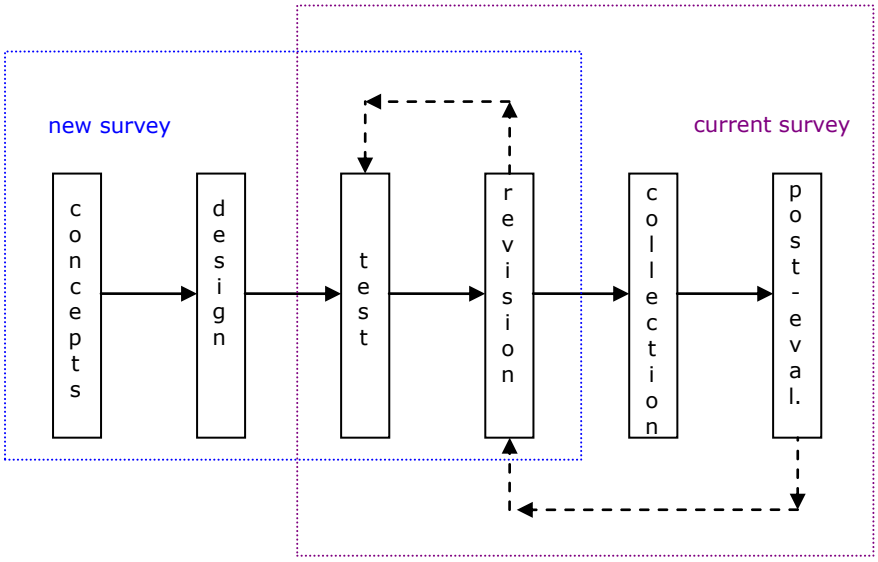
Principle B.4.3. Questionnaire testing strategy
New questionnaires or new questions/sections of pre-existing questionnaires should be evaluated through a testing strategy. Questionnaires already in use in current surveys should be periodically evaluated. The testing or evaluation strategy should include both pre-field and field testing.

Guidelines

The European Statistics Code of Practice requires that all questionnaires are systematically tested prior to data collection. When designing a new questionnaire or redesigning an existing one, it is necessary to define an overall strategy for development and testing, taking into account the survey objectives, the population under investigation, and the available resources in terms of budget, personnel and time. The strategy should follow different approaches depending on whether the survey is completely new, or a questionnaire already used in a current survey is to be revised, in which case it is possible to take advantage of data from previous occasions.

In general, a test strategy should include the application of mixed methods, as each one is oriented to solve different and complementary problems. For new questionnaires at least one pre-field and one field test should be planned, as well as the possibility to make appropriate modifications to the instrument and re-submit it for evaluation. In statistical literature, it is common to distinguish between pre-field methods, whereby some or all the conditions in which the tests are conducted may differ from the survey actual situation, and the so-called field methods, which are carried out in conditions that resemble the real survey as much as possible (e.g. settings, data collection technique, length of the questionnaire, order of questions). In current surveys, the possibility of revising the questionnaire is limited by the requirement of data comparability over time. However, testing should be performed periodically, exploiting the data collected on previous survey occasions (post-evaluation), and using the most appropriate methods to test previously developed versions of the questionnaire, as represented by the following figure.

In the case of a global redesign, where important innovations are required also in the questionnaire, the



steps concerning questionnaire design should be run through again, as in the case of new surveys. The existence of questionnaires used on previous survey occasions should not impede the improvement process.

Questionnaire design

Before designing a questionnaire, a study of literature and an analysis of existing surveys on the topic or on related topics should be carried out. Survey objectives should be translated into clear questions, allowing to derive the variables and indicators of interest.

In the preliminary stages of questionnaire design, when the questions are not yet well defined, it is recommended to conduct focus groups and qualitative interviews to see how concepts and issues are perceived and interpreted by potential respondents.

In preparing the questionnaire, some general principles should be followed. Some key elements should be clearly visible at the beginning of the questionnaire, such as the institution in charge of the survey, the survey title and topic, explanations about survey objectives, the request for cooperation and assurance of confidentiality. The inclusion of each question should be carefully evaluated against respondent burden. The question whether to provide a final text-box for additional comments by respondents should be evaluated. A sentence on appreciation for respondent participation should end the questionnaire.

The questionnaire is one of the main sources of measurement error and item nonresponse. Therefore, in designing survey forms, particular attention should be paid to the prevention of such errors.

Researcher should be aware that the administration of a questionnaire induces a cognitive process in the respondent, consisting of several phases (encoding, comprehension, retrieval, judgment, reporting). The analysis of these phases in a cognitive perspective helps to identify possible sources of error and allows to better evaluate several issues, including the choice of the reference period, the use of proxy respondents, the formulation and sequence of questions and response options. In general, the main topic of the survey should be phased in during the interview and the language should be as neutral as possible, consistently using concepts and terms throughout the entire questionnaire. The answer categories should be mutually exclusive, exhaustive or provide for the possibility of the category

"other". Number and order of the options to be considered should be assessed in relation to the technique. The choice of open-ended versus closed questions should be carefully evaluated considering the type of information required, the level of accuracy sought, respondent characteristics, the resources available for data entry and coding tasks. When respondents are highly qualified in the field of interest, such as in many cases in business surveys, the language should be technical.

In surveys where there is a questionnaire common to several Countries, harmonisation of variables and accurate translation process should be pursued by using sound methodologies.

With regard to the questionnaire structure, the sections of the questionnaire and - within the sections - the questions, should follow a logical order, understandable to respondents.

Particular care should be taken in the preparation of instructions for respondents or interviewer. The instructions should be clear and easily accessible.

Questionnaires are not only a means of gathering information but also a communication tool. They should appear attractive and professional at the same time. The graphic solutions adopted in questionnaires should be uniform and allow to clearly identify the different types of text associated with: titles of sections, questions, answer categories, instructions, routing, and so on.

Questions on topics that respondents may perceive as embarrassing or highly sensitive may lead to inaccurate answers. To overcome this drawback, it is appropriate to carefully evaluate several factors such as: to adopt a minimally intrusive technique, i.e. preferring a telephone survey or a self-administered questionnaire to a face-to-face technique; to introduce the sensitive topic in a gradual manner, to provide additional assurances of confidentiality; to make further attention to the neutrality of the language used; to explicitly permit respondents not to answer the question; to evaluate, whether interviewer characteristics may have an influence on responses and to establish specific criteria for their selection; to promote specific training of interviewers.

Questionnaire testing

The drafting of questionnaires should be evaluated by means of a set of pre-field and field tests. The aim is to explore a variety of issues ranging from graphical layout, to language, instructions for questionnaire routing, order of sections and questions, usability of questionnaires. Preliminary assessments as informal tests or experts' reviews may also help identify major defects in the instrument. Cognitive interviews are generally oriented to assess whether respondents understand the questions in the manner sought by researchers and if they may remember and provide answers to the questions. Generally, though not necessarily, cognitive interviews are carried out in very different conditions from those that actually occur during the survey. For example they are conducted in the laboratory and with test-persons, and it is therefore necessary to integrate these methods with others carried out in the actual survey conditions, or in situations as close as possible to the real ones.

Many tests are carried out directly on the field, such as the observation of behaviour of respondent, interviewer or both using structured schemes (*behaviour coding*). The aim is to understand whether interviewers plays their role properly, how respondents reacts, but also their interaction. Other applicable methods are respondent and interviewer debriefings and intense or follow-up re-interviews.

Interviewers should be involved in the evaluation process since they may provide important feedback on questionnaire adequacy (interviewer debriefing).

It is advisable to include questionnaires among the tools that are tested by means of pilot surveys, if carried out, since this allows for the evaluation of the instrument in conditions similar to those of the survey.

In current surveys, specific analyses aimed at deriving information on the quality of the instrument used and at implementing improvements for subsequent survey occasions should be carried out. In particular, *ex-post* indirect assessments should be conducted, based on the analysis of item nonresponses and the frequencies of "do not knows", "not remember" and other similar categories to some questions. Also the frequency of edit failures in the editing and imputation procedure may be a signal of potential problems in the questionnaire.

When developing electronic questionnaires, *ad hoc* software should be used. The software should be agreed as an internal or international standard. In such a situation, the electronic version of the questionnaire with respect to its functionality and, where possible, with respect to its usability, should be tested.

Electronic questionnaires permit to introduce quality controls on the data collected during the interview. The extent of such quality controls should be carefully evaluated and properly balanced not to overload the interview with frequent interruptions that could compromise its completion. Routing errors should be treated by means of "hard" controls (i.e. not allowing the continuation of the interview if they are not resolved). Range errors may be treated in a "hard" mode if the defined domains are large or in "soft" mode within sub-domains. Finally, consistency errors should be controlled in "hard" or "soft" mode depending on the importance of the variables they apply to.

The factors guiding questionnaire design, as well as the results of the testing phase, should be properly documented.

Some references

Brancato G., Macchia S., Murgia M., Signore M., Simeoni G., Blande K., Körner T., Nimmergut A., Lima P., Paulino R., Hoffmeyer-Zlotnik J.H.P. (2005) Handbook of Recommended practices for questionnaire development and testing in the European Statistical System, Eurostat
http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/RPSQDET27062006.pdf

C. Survey frames

C.1. Frame updates and coverage

Principle C.1. Frame updates and coverage
The frame should properly cover the target population. It should include accurate and updated information to be used to contact survey units. In the case of sample surveys, where possible, the frame should include information so as to allow the selection of an efficient sample (e.g. stratification or selection of units with probabilities proportional to some measure of size).

Guidelines

Both the target population and the frame to be used to identify and contact survey units should be clearly defined. If the inadequacy of the frame implies redefining the target population (i.e. the so called frame population is investigated instead of the target one), such decision should be documented and made known to users.

If a frame has been chosen among multiple alternative frames, this choice should be justified. The production of a frame as a combination of alternative frames should be based on sound and generally accepted methodologies and procedures (see section E.4).

The frame should be as up-to-date as possible with respect to the reference period of the survey.

The frame coverage should be mainly evaluated with respect to undercoverage. Undercoverage, in fact, may lead to substantial bias in survey estimates. If the phenomenon is not ignorable, actions aimed at improving the frame should be performed before using it for selecting the sample and contacting the units. If, despite the adoption of preventive actions, the size of undercoverage remains significant, it is appropriate to try to compensate for potential bias during the production of the estimates, using *ad hoc* methodologies (see section E.5).

Auxiliary information available in the frame may be used for the design and selection of the sample. Errors in the variables used in the design or selection of the sample may lead to loss of precision of final estimates. If the incidence of such errors is not negligible, the opportunity of using procedures for their identification and correction should be considered.

Not up-to-date or inaccurate information on the units in the frame may prevent contact, with consequent reduction of the sample size with respect to what planned, and loss of accuracy. If the phenomenon is not negligible, the risk of bias in final estimates may also occur.

Frame characteristics and actions carried out on it should be documented.

Some references

Statistics Canada (2009) *Statistics Canada Quality Guidelines Fifth Edition – October 2009,* Catalogue no. 12-539-X, Ottawa. http://www.statcan.gc.ca/pub/12-539-x/12-539-x2009001-eng.pdf

D. Data collection

D.1. Use of administrative data

Principle D.1. Use of administrative data
Permanent partnerships with providers should be established to assure timely acquisition of administrative records with a good level of accuracy. Variations on both definitional and technical aspects should be agreed upon and check and monitoring of provisions should be carried out.

Guidelines

Whenever data are collected using administrative sources, having a say in the legal aspects that regulate the production of administrative data represents a favourable condition to ensure the relevance of administrative data for statistical purposes and the comparability of statistical products over time.

Good relationships with the agencies providing administrative archives should be established and maintained, by setting up formal agreements. These agreements should cover procedures and timing for data transmission, required quality level of administrative archives, documentation supporting the archive transmission, and also statistical information feedback to archive providers. The objective is the improvement of the production process and, in general, of data quality.

It is necessary to fully understand the context in which administrative data are generated and managed, i.e. relevant legislation, purposes and uses of administrative archives.
Indeed, such information has a considerable influence on the use of administrative records for statistical purposes (for example with respect to: coverage, content, concepts and definitions used, frequency and timeliness of administrative records, quality of keyed information, stability over time). In particular, all metadata necessary for the proper use of administrative archives should be requested to the Institution in charge of them. Furthermore, it would be appropriate also to collect information on the underlying quality of these archives, and on whether and what procedures have been applied to improve the completeness and consistency of the information they contain.

The evaluation of the quality of the administrative archive through the evaluation of coverage with respect to the target population for statistical purposes (eligible and non eligible units for statistical purposes, undercoverage) should be performed before starting the data processing phase. Eligible units should be analysed to evaluate the proportion of missing or inconsistent information. To this end, it is advisable to calculate a set of quality indicators or to perform more complex analysis based on integrations and comparisons with other archives (see Section E.4). The impact of potential errors in administrative sources on final data quality should be evaluated.

The transmission of administrative archives should be carried out by using secure channels and Istat standard–compliant protocols. It is advisable to use transmission checks to prevent data loss.

All phases of acquisition and processing of administrative archives should be documented.

Some references

Statistics Canada (2009) *Statistics Canada Quality Guidelines Fifth Edition – October 2009,* Catalogue no. 12-539-X, Ottawa. http://www.statcan.gc.ca/pub/12-539-x/12-539-x2009001-eng.pdf

D.2. Direct data collection

Principle D.2.1. Choice of the data collection mode
Data collection techniques and instruments should be selected taking into account survey objectives and target population characteristics, and in order to maximise data quality while limiting statistical respondent burden and costs.

Guidelines

The data collection mode should reduce statistical burden and costs and, at the same time, maximise data timeliness and accuracy. The selection should take into account many different factors, sometimes difficult to combine. In particular, it is convenient to analyse the following aspects.

It should be considered whether the general objectives and the specific topics of the survey address towards the choice of a particular technique. For example, when collecting data on sensitive topics it is advisable to use a self-administered mode. Also telephone interview, in which the interviewer is present but less intrusive, may be used, if overall convenient. If a face-to-face interview is chosen, despite dealing with sensitive topics, strategies to protect the respondent privacy should be adopted, e.g. closing answers in a sealed envelope or using the method of randomised responses.

The complexity and extent of the topics to be investigated should be taken into account. A topic that requires a high articulated interview, i.e. a questionnaire with many "skips" and complex "routing", may not be surveyed by a self-administered paper questionnaire, while it may be more effectively managed through the use of computer-assisted techniques, such as computer assisted face-to-face (CAPI) or telephone (CATI) interviewing. In telephone surveys, the interview length should be limited, compared to mail and face-to-face techniques.

It is useful to analyse the target population and to ascertain the existence of characteristics which are positively related to survey participation in order to choose the data collection mode which is more appropriate to meet respondents' needs. Mixed modes should be preferred when they help meeting manifold needs. How the information of interest is organised should be taken into account for surveys on businesses or institutions. For example, mail or web modes should be preferred when information should be retrieved from archives, from a specific office of the business surveyed or outside the business (e.g. from an accountant). For surveys on public bodies and institutions, data collection modes should be agreed upon via formalised agreements.

It should be evaluated which is the best period of the year to carry out the survey. In the case of surveys on public bodies and institutions, the data collection period should be agreed with them. During the data collection period the best time to contact, inform, schedule an appointment and finally visit respondents, should be chosen. Such time constraints have an impact on the organisation of the data collection phase. Some techniques are more suitable than others to meet the needs of respondents (for example, as households are easier to contact in the evening, then it becomes preferable to carry out telephone interviews than visits).

The best technologies available should be utilised. If possible, computer-assisted techniques should be preferred, as they ensure: efficiency gains in the survey production process resulting in improved timeliness; anticipation of checks (consistency, domain and flow) on responses provided in the data

collection phase with opportunity to verify responses during the interview; overall cost reduction and, in particular, reduced use of paper in compliance with national policies.

In the case of adoption of mixed modes, the "mode effect", that is the impact in terms of variability and bias attributable to the use of different techniques on subsets of the target population, should be evaluated, also experimentally.

Principle D.2.2. Use of external intermediate bodies or data collection contractors
In order to reduce non-response and achieve a high quality level of the information collected, adequate measures should be taken to improve cooperation with intermediate bodies involved in data collection. Actions to monitor the activities of these bodies should be planned.

Principle D.2.3. Respondents
In order to ensure adequate response rates and a high quality level of the information collected, participation of respondents should be encouraged, and measures on the collection phase should be produced and evaluated.

Principle D.2.4. Interviewers
Interviewers should be selected, trained and monitored so as to ensure that the data collected is as accurate as possible. It is good practice that the interviewer bias effect be assessed through *ad hoc* studies.

Guidelines

In direct data collection, statistical information is collected from (all or part of) the units of a population, which may also consist of administrative bodies, according to a statistical survey design.

In some cases, it is possible that the support of intermediate bodies, such as municipalities, chambers of commerce, and also of private companies is used in data collection operations. In this case, before the beginning of the data collection phase, it is necessary to clearly establish responsibilities, agree on the timing and activities to be performed, preferably with formal documents, and maintain good relationships with these intermediate bodies. It is also necessary to organise training meetings and to develop a support and supervision system. During data collection, it is necessary to monitor the activities of intermediate bodies also through the analysis of *ad hoc* indicators, to agree upon data file transmission, even if partial, and to check it.

Since one of the most critical aspects during data collection concerns unit participation, a series of actions to prevent unit nonresponse should be planned and implemented. In fact, unit nonresponse may lead to not negligible bias in the estimates. Examples of actions to prevent nonresponses are: sending a presentation letter signed by the President or the Central Director; providing respondents with a brief description of the objectives of the survey, explicitly guaranteeing the protection of confidentiality, establishing a toll-free line or an email for respondents. If, despite prevention, it is believed that there might be a problem of nonresponse bias, it is advisable to try to compensate for it through the use of appropriate methodologies when processing data (see section E.5).

Especially in some contexts, for example in business surveys, it should be carefully evaluated who is the most appropriate person to fill in the questionnaire and how he/she may be effectively contacted. Similarly, the aspect concerning the organisation of the information required from respondents and its accessibility should be deepened. In some circumstances, responses from proxy respondents are also

accepted, namely from individuals other than the person from whom information is intended to be collected, such as parents who are asked to answer questions relating to their children, too young to do it themselves. Although this choice is sometimes necessary and helps in keeping the nonresponse rate low, it should, however, be borne in mind that the answers provided by proxy respondents may differ systematically from those reported directly by respondents. The benefits and risks for data quality, resulting from the use of proxy respondents, should thus be carefully evaluated.

Other aspects to be taken into account are: the choice of the best time to call or visit the survey unit which implies careful planning of the attempts to contact it and the follow-up strategy; the implementation of a monitoring system based on indicators of nonresponse, following the standards of the Institute. The definition of the minimum set of information allowing for a unit to be considered respondent is a pre-requisite for the development of such a system.

During data collection, the progress of the survey should be constantly monitored according to the developed tools, and all the necessary actions to achieve the quality targets should be implemented. Decisions should be taken on the basis of objective and standard indicators, such as those defined in SIDI/SIQual[4], or more detailed ones, but always ensuring consistency with the standards.

Interviewers are essential to the success of data collection operations. Some characteristics of interviewers and their level of training may influence the quality of the responses collected. If the socio-demographic characteristics of interviewers are related to the information of interest, it is necessary to establish criteria and requirements for their selection. In addition, interviewers should receive extensive training on all aspects concerning survey objectives and questionnaire contents, communication and contact with potential respondents, ways to convert refusals, management of questionnaire skips, use of electronic questionnaire, etc. Interviewers should be provided with all the useful material and with a manual of instructions and other informative material about the survey. During the data collection phase, frequent consultations with interviewers (debriefing) should also be organised to bring out any problems found and to have the opportunity to timely find solutions.

Monitoring tools allowing for in-progress monitoring of interviewer work should be designed and implemented. They range from field supervision, to follow-up calls and analysis of performance and quality indicators, also applying appropriate methods for summarising them (for example, control charts). The interviewers' workload should be monitored and excessive turnover should be avoided. The risk of a potential interviewer bias effect on data should be considered in advance and the fieldwork should be possibly organised in such a way to allow evaluating it with specific statistical measures.

Data transmission from intermediate bodies, private contractors or directly from respondents, should be carried out using secure transmission channels and Istat standard–compliant protocols.

The data collection phase should be properly documented.

[4] Istat documents currently its statistical production processes and their quality in an information system on quality of statistical production processes, called SIDI. A large part of this information is available for users on the website of the Institute via the navigation system called SIQual.

Some references

Statistics Canada (2009) *Statistics Canada Quality Guidelines Fifth Edition – October 2009,* Catalogue no. 12-539-X, Ottawa. http://www.statcan.gc.ca/pub/12-539-x/12-539-x2009001-eng.pdf

Statistics Canada (2010) *Survey Methods and Practices.* Statistics Canada, Catalogue no. 12-587-X, Ottawa. http://www.statcan.gc.ca/pub/12-587-x/12-587-x2003001-eng.htm

E. Data processing

E.1. Coding

Principle E.1. Coding
The coding process (automated, computer assisted or manual) should ensure a high level of quality. Quality measures should be collected and evaluated.

Guidelines

Coding is the technical procedure for converting textual information into numeric codes. The coding process may be defined as:

- automated, if it is performed by a software application that assigns, in an automated way, codes to textual information;
- computer assisted, if the coding process is performed by respondents, interviewers or coders assisted by a specific software application;
- manual, if the coding process is performed by *ad hoc* trained operators (coders) without the help of a specific software application.

In case of automated or assisted coding, some complex cases often remain uncoded; such cases should be resolved using expert coders.

Coding errors should be limited as much as possible and specific actions should be taken to prevent them, such as, for instance, providing coders with a specific software application to assist them and with appropriate training.

The use of generalised software designed for assisted and automated coding is recommended in order to ensure the application of efficient and sound methodologies for coding operations.

Collection of quality measures is required in order to evaluate the coding process. For this purpose, textual information should be converted into electronic format. Several methods may be used to evaluate coding quality, as recoding of a sample of items carried out by a team of expert coders. In case of automated coding, checks may be based on the computation of process indicators such as the recall rate indicator that is the percentage of items coded by the procedure.

If coding operations are carried out by external contractors, the coded data should be transmitted to the Institute by using secure transmission protocols. The transmission should be equipped with documentation, defined in the contract, needed to evaluate coding activity and quality.

The coding process should be evaluated as a whole taking into account the trade-off between accuracy and the time required to finalise the operations that should not lead to high delay in data dissemination.

Evaluation of coding errors and of the time required for their treatment may point out weaknesses in the process and allow for the definition of related improvement actions to be implemented on subsequent survey occasions.

The analysis of coding errors and of uncoded text information may point out elements to classification experts to improve the information base of coding systems.

Some references

Istat (2007). Metodi e software per la codifica automatica e assistita dei dati. *Tecniche e strumenti*, n. 4, 2007

E.2. Data capture

Principle E.2. Data capture
The data capture process performed either by operators or using optical character recognition techniques (OCR) should ensure high quality level. Quality measures should be collected and evaluated.

Guidelines

The operators that are in charge of converting paper forms into electronic format (keyers) should be properly trained and provided with adequate tools.

The software used for data entry should include a range of controls to minimise registration errors: "hard" controls on identification codes and preferably "soft" ones (domain, flow and consistency) on other data. Controls, however, should not be excessive, in order to avoid too frequent interruptions in the registration.

It is preferable to use generalised software designed for data capture as it allows to manage the controls to be performed in an efficient manner.

Concerning the adoption of OCR techniques, part of data often may not be captured automatically, because some characters may not be recognized by the software, or because some questionnaires may be in bad conditions. Consequently, OCR techniques should be combined with registration by operators.

In case of data capture carried out by subcontractors, the recorded data should be sent to the Institute by using a secure transmission method; evaluation of reports on the recording quality should be carried out. Such reports might be drawn up internally in the control phase of the data received, or by the subcontractors if foreseen by contractual agreement. Quality measures of the data entry should be collected also when data capture is performed internally.

The quality of the data capture process should be evaluated both in relation to accuracy (minimising the incidence of errors in recording) and in relation to the time spent for that phase, which should not cause a large delay in the data release stage.

Assessments made on accuracy and required time of data capture may be used to improve the production process on subsequent survey occasions.

E.3. Error detection and treatment

Principle E.3.1. Designing a strategy for error detection and treatment and item nonresponse in collected data
The overall procedure for error detection and treatment and item nonresponse should be designed taking into account the following issues: survey characteristics; amount and type of data to be checked; data timeliness; available methods, auxiliary information and resources. Furthermore, it should be organised in phases according to the different error types and to the proper methods for correcting them.

Principle E.3.2. Interactive review
The clerical review should be based on a transparent, reproducible and documented system of error detection and treatment. It should be efficient with regard to the set of checks to be done and to reviewers monitoring.

Principle E.3.3. Editing and imputation
The editing and imputation procedure should be based on sound statistical methodologies and should be the most appropriate for the survey data; it should be documented and assessable. If available, generalised software implementing these methodologies should be used.

Principle E.3.4. Ex-post evaluation and documentation of the editing and imputation phase
The editing and imputation procedure and its impact on data should be periodically evaluated and documented. The assessment results should be used to improve it and, possibly, other phases of the survey process.

Guidelines

All activities related to error detection and treatment should be made explicit in an overall strategy that takes into account survey characteristics, type and amount of collected data and expected errors. The resources required for tool development and testing as well as for their application during the survey should be also assessed for each step of the strategy. These activities should be designed taking into account the other sub-processes of the statistical production process.

The procedures for error detection and treatment should minimise changes in data, i.e. consistent data should be obtained by changing the collected data as little as possible.

Edit rules should be the result of collaboration among subject matter experts, survey staff and experts on editing and imputation methodologies. If available, information from previous survey occasions should be taken into account in defining such rules. Edit rules should be coherent and not redundant, in order to avoid excessive data correction (*over-editing*). The editing and imputation strategy should be organised by priority, allocating more resources on the treatment of most serious errors and most important units and variables. Anyway, the interactive review should be limited to the most relevant errors that may not be solved automatically. In order to detect such situations, selective editing and methods for the identification of outliers and influential observations may be used. The different stages of the editing and imputation strategy should be periodically evaluated, also by means of simulations or experiments, to either confirm their validity or make the necessary changes.

The involvement of reviewers in the procedures for error detection and treatment may cause an impact in terms of bias and increased variability of estimates. To limit this impact, especially when manual imputation is planned, reviewers should be properly trained and equipped with appropriate written

guidelines. The guidelines should explain the rules to be followed when implementing edits and treating different error types. The guidelines should be developed, tested, and then reviewed periodically, and their implementation should be monitored in order to avoid "creative-editing". A system for supporting and monitoring reviewers should be also set up. In case of potential reviewers' effect on estimates, it is recommended to make an assessment also by means of experiments.

Editing and imputation methods and tools should be chosen taking into account statistical literature, available experiences on similar data, specific guidelines, standards or recommendations developed at national or international level. Finally, they should be based on sound methodologies.

For error detection, different methods and tools should be applied according to the error type. Missing values should be distinguishable from undue values and, in case of quantitative variables, from structural zeroes. The definition of deterministic rules to detect systematic errors should be the result of the analysis of indicators on edit rules. Systematic errors should be detected and corrected before random errors and before selective editing. As already mentioned, the detection of influential errors follows an approach based on selective editing, whose priorities should reflect a score function evaluating the risk of error and the influence on estimates. The identification of influential units may also be performed by the analysis of the main preliminary estimates of the survey (macro-editing). For outlier detection robust methods, ranging from simple univariate analysis to complex graphical methods based on the relationships between variables in different subpopulations, should be used. In any case, the plausibility of outliers should be carefully evaluated before correcting it. Finally, for random errors, generalised software implementing a sound methodology, such as the Fellegi-Holt paradigm (minimum change principle) should be used.

Imputation consists of replacing missing data or values flagged as erroneous during editing with plausible values. A good imputation procedure should be objective, reproducible, assessable; it should make efficient use of available auxiliary information and guarantee internal consistency of imputed records. Imputation methods may be classified as deterministic or stochastic. The former include deductive imputation, imputation from time series, imputation with the average value, imputation based on regression model without stochastic component and nearest neighbour imputation. The latter include random donor-based imputation, nearest neighbour imputation when the donor is randomly selected among a set of candidate units, imputation based on regression models with stochastic component and other deterministic methods to which random residuals are added. Deductive imputation should be the first method taken into account when available information leads to a unique acceptable value and when the error nature is well known, as in the case of systematic errors. Any imputation method corresponds to assume, implicitly or explicitly, a model based on auxiliary information. The selection of auxiliary variables should be made taking into account the strength of their relationships with variables to be imputed and their contribution to explain the non-response mechanism. An imputation model with auxiliary variables should be carefully validated for each variable to be imputed as well as for groups of variables. Another critical aspect is the choice of the donor: a given donor should be used for a limited number of recipients, while for a specific recipient the number of different donors should be limited.

With regard to the application of editing and imputation procedures, generalised software implementing sound methodologies should be used whenever available. The editing and imputation system adopted should be flexible enough to allow for the introduction of changes or integrations with limited costs.

To ensure that the different sub-phases of the editing and imputation process are assessable, it is necessary to keep both original and imputed values at the various stages of the procedure. Indirect evaluation of the different steps of the procedure should then be performed by calculating indicators on the amount of changes made (SIDI/SIQual standard indicators) and indicators on the changes in the distributions of the variables of interest.

When editing and imputation techniques are applied, the estimation of the additional variability due to imputation should be taken into account.

Information derived from editing and imputation procedures, such as, for example, the frequency of activation of edit rules or the imputation rate per variable, represents the alarm bell of potential problems in earlier stages of the production process (for example due to shortcomings in the questionnaire) and may give insight on the main sources of error. This worthwhile information should be analysed and used to improve subsequent survey occasions.

All the steps of the strategy for error detection and treatment in collected data should be properly documented.

Some references

Luzi O., Di Zio M., Gurnera U., Manzari A., De Waal T., Pannekoek J., Hoogland J., Tempelman C., Hulliger B., Kilchmann D. (2008) Recommended Practices for Editing and Imputation in Cross-Sectional Business Surveys. EDIMBUS project http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/RPM_EDIMBUS.pdf

E.4. Integration of data sources

Principle E.4. Integration of data sources
Integration of sources should be carried out in accordance with the objectives of the analysis and should be based on sound and generally approved methodologies. The integration process should be clearly defined and each step should be tested. The goodness of result of the integration process should be evaluated by computing appropriate indicators. The integration process should be carried out abiding by the regulations concerning confidentiality.

Guidelines

Integration of different data sources may have different purposes. Typically, in the case of administrative records, integration is carried out to address coverage problems of an archive, to make new variables available in the reference archive or to impute missing values. Integration of survey data with administrative data may be carried out to overcome any lack of response (complete or partial), to make new variables available or to carry out "records check" studies aimed at identifying and assessing the impact of potential measurement errors.

Integration between administrative sources or between administrative sources and survey data may be done in different ways. If the units have a unique error-free identification code, then it is possible to integrate data via exact matching (merging) based on the identification code. In the absence of an identification code, if key variables that jointly considered may contribute to identify the unit (e.g. name, date of birth, address ...) exist, then matching may be conducted through "record linkage" procedures. Record linkage is deterministic when it is based on formal rules to determine whether pairs of records in two separate sources refer to the same unit. It is probabilistic when the decision rule is based on probabilistic criteria.

Often integration procedures consist of a combination of different integration methods. In this case it is necessary that the entire procedure is well defined, clearly specifying the order of the various methods.

The methods used in the integration process should be sound and used at international level. All processing tasks performed on origin data sources for integration purposes should be documented.

Integration of different data sources should be carried out by means of generalised software. The development of *ad hoc* software codes should be followed by a thorough testing phase to prevent programming errors that may affect the accuracy of integration process results.

The quality of the entire integration process should be assessed by calculating proper indicators. Whenever possible in record linkage applications the false match rate (falsely matched records that in fact are two separate units) and the false non-match rate (units erroneously not matched by the procedure) should be estimated.

All integration tasks should be performed in accordance with the regulations on confidentiality.

The whole integration process should be documented.

Some references

Eurostat (2008) "State of the art on statistical methodologies for integration of surveys and administrative data". *Report of the WP1 of the ESSnet Statistical Methodology Project on Integration of Survey and Administrative Data*.
http://www.cros-portal.eu/content/wp1-state-art

E.5. Estimation

Principle E.5. Estimation
In order to produce estimates, the data collected and appropriately treated (coding, editing, ...) should be processed according to consolidated and approved methodologies (at national or international level). The use of auxiliary information for developing model-based estimates should be justified, assumptions underlying the models should be made explicit and their actual validity should be assessed. The produced estimates should be released together with estimates of sampling errors to allow for their correct use and interpretation.

Guidelines

The procedure to derive estimates (totals, proportions, contingency tables) should be well and clearly defined.

Typically, in case of surveys based on probability samples, estimates are calculated using weights that are derived directly from the sampling design (sampling weights), then corrected to compensate for the impact of non-sampling errors (unit nonresponse, undercoverage) or to exploit the auxiliary information available in order to obtain more accurate estimates (e.g. calibration).

In sample surveys that make use of non-probability samples, estimates are generally derived by using appropriate statistical models. It should be noted that the use of models may also occur in surveys based on probability samples. In this case sampling weights are usually not taken into consideration. In general, the use of model-based estimation procedures should be justified and the related assumptions should be plausible and, if possible, tested. In some cases, model validations may be carried out *ex post*, where new surveys, based on probability samples, are carried out on the same population.

If auxiliary variables are used in the estimation process, they should be correlated to the survey variables and updated. The use of a particular model should be documented as well as the main underlying assumptions. If several auxiliary variables are available, it should be explained how the variables actually used were selected.

Estimates should be released with error measures. These measures should take into account, where possible, the main errors (sampling and nonsampling) observed in the whole process.

The criteria for estimate dissemination should be established prior to their production: that is, by fixing the error level above which estimates are not disseminated.

The estimation procedure should be carried out using generalised software. The use of *ad hoc* software should be restricted to peculiar cases, and it should be preliminarily and accurately tested before producing final estimates.

All results of estimation processes should be replicable (exactly or with minor approximations), meaning that by repeating all processing procedures, the same results have to be obtained.

Probability sample surveys

A weight is usually associated to each sampled unit; the weight may be the direct sample weight (obtained as the inverse of the inclusion probability) or the same corrected to compensate the impact of nonsampling errors (unit nonresponse, undercoverage) or to exploit available auxiliary information to derive more accurate estimates of the values of interest (e.g. calibration).The procedure to derive estimates using weights should follow sound and generally accepted methods and techniques and should be transparent and documented.

Correction of weights to compensate nonsampling errors (unit nonresponse, undercoverage) should be carried out using sound methodologies, that are both nationally and internationally accepted. The auxiliary variables used in the corrections of weights should be explanatory of the problem that needs to be corrected and updated. Auxiliary variables used to correct weights in order to improve estimate accuracy should be related to the variables for which it is necessary to produce estimates.

An estimate of sampling variance for the most important estimates should be produced both for the entire population and for the main domains. Such an estimate needs to take into account the design characteristics (stratification, multistage selection, etc.) and the corrections made to weights. When the estimate of sampling variance is obtained by approximate methods, such choice should be documented.

The impact on estimates in terms of variance and of bias due to non-sampling errors should be considered, if possible.

Some references

OMB (2006) Standards and Guidelines for Statistical Surveys. Office for Management and Budget, The White House, Washington, USA.
http://www.whitehouse.gov/sites/default/files/omb/assets/omb/inforeg/statpolicy/standards_stat_surveys.pdf

Särndal C.E., Lundström S. (2005) *Estimation in Surveys with Nonresponse*. Wiley, New York.

Statistics Canada (2009) *Statistics Canada Quality Guidelines Fifth Edition – October 2009,* Catalogue no. 12-539-X, Ottawa.
http://www.statcan.gc.ca/pub/12-539-x/12-539-x2009001-eng.pdf

Statistics Canada (2010) *Survey Methods and Practices.* Statistics Canada, Catalogue no. 12-587-X, Ottawa.
http://www.statcan.gc.ca/pub/12-587-x/12-587-x2003001-eng.htm

E.6. Seasonal adjustment

Principle E.6. Seasonal adjustment
Seasonal adjustment procedures should be aimed at removing the seasonal component of time series. Seasonally adjusted data should not have residual seasonal effects. The approach used to seasonal adjustment of data should be explained and based on sound and generally accepted methodologies. The assumptions underlying this approach should be periodically reviewed. Users should be clearly informed about the release of seasonally adjusted data.

Guidelines

Seasonal adjustment procedures should be performed only when there is clear statistical evidence and economic interpretation of the seasonal/calendar effects. A time series should be seasonally adjusted when data exhibit seasonal patterns and when the underlying seasonability may be properly identified, that is when it is not dimmed or hidden by a high level of irregular variations.

Seasonal adjustment should be preceded by preliminary treatment of data, aimed at correcting the influence due to the different number of working days, holidays (fixed or mobile, civil or religious), to anomalous values (outliers) and, finally, to exceptional events (strikes, disasters, etc..). All the procedures for preliminary treatment should follow sound and generally accepted methodologies and should be adequately documented.

The estimate of the seasonal component should be carried out by using sound and generally accepted procedures. It is necessary to regularly revise the specifications used for pre-treatment and to estimate the seasonal component in order to take into account any revisions of the raw data already released and the dissemination of new data.

The methodology used should be properly documented, along with the software used and its version. Specifications of the procedure used should be released on request to external users.

It is necessary to use standard diagnostics (graphs, absence of residual seasonality, stability of the seasonal component, model residuals diagnostic ...) that are included in the output of the procedure used to validate seasonal adjustment.

Some references

Eurostat (2009) "ESS Guidelines on Seasonal Adjustment. 2009 Edition". *Eurostat Methodologies and Working Papers*, Luxembourg.
http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-09-006/EN/KS-RA-09-006-EN.PDF

E.7. Revision policy

Principle E.7. Revision policy
Revision policy should report modalities and revision timing of estimates. The revision procedure should be clearly defined and made known to data users. Revision studies and analyses should be carried out on a regular basis and should be used in restructuring the production process.

Guidelines

In some cases, the need to disseminate timely estimates comprises the release of preliminary or provisional estimates, that are later revised as new (or updated) information becomes available. Sometimes such revisions may also be determined by the application of different estimation procedures, changes in methodologies or unexpected events.

All information regarding the process for revising estimates should be clearly and explicitly specified in the revision policy, pointing out the sources used and their timeliness, the planned number of revisions, their reasons and calendar. The revision policy should be communicated in advance to data users. In publishing estimates subject to revision, users should be informed, presenting estimates with information on revision timing and modalities.

Each revision should be documented by using "revision triangles"[5], that allow for the reconstruction of the history of the released estimates and for an assessment of the impact of the revision policy. The documentation should also include the results concerning the computation of the main revision indicators and provide references to documents with more detailed analyses of revisions (if available). Revision triangles should be regularly updated with the dissemination of new data. The choice of the information disseminated in the triangle (level and/or variation data, seasonally adjusted or not) should take into account the requirements of external users.

If the revision analysis revealed a systematic trend of estimates (tendency of preliminary estimates to underestimate or overestimate phenomena), efforts should be done to find the causes and, therefore, act on the production process in order to remove them, if possible.

Occasional revisions, not included in the revision policy, should be documented and justified. Users should be informed of such revisions and of the reasons why they were made.

Some references

Istat (2010) L'analisi delle revisioni delle informazioni statistiche congiunturali. *Approfondimenti*. http://www3.istat.it/salastampa/comunicati/non_calendario/20101013_00/approfondimenti_analisi_revisioni.pdf

[5] Istat releases the revision triangles along with some economic indicators and the quarterly main aggregates of National accounts.

E.8. Data validation

Principle E.8. Data validation
Prior to dissemination survey results should be evaluated together with subject matter experts to determine whether or not there are anomalies.
Where practicable, results should be compared with the same results obtained on previous occasions of the process or with similar results obtained by other processes within the same organisation or from outside agencies.
Moreover, quality indicators of the process (including, for sample surveys, the level of precision of final estimates) should be calculated and analysed, also to assess the possible introduction of improvement actions on subsequent survey occasions.

Guidelines

Before releasing survey results, they should be evaluated through comparisons with results from previous occasions of the same survey and through comparisons with external sources.
Such sources may be internal statistical sources, sources external to the Institute or administrative sources. Eventual differences should be justified and documented.

If possible, consistency of results with respect to ratios that may be considered almost constant or subject to minor changes in the short period should be controlled (for example some demographic ratios). Again, eventual differences should be justified and documented.

In case of suspicious values, the results should be evaluated prior to their release by subject matter experts from the Institute or by external experts such as academics or business associations. If the evaluation is performed by external experts, data confidentiality should be assured. In any case it is preferable to involve internal or external experts in the validation that are not directly involved in the data production process.

During the validation phase, quality indicators[6] such as, for example, coverage error rate, response rate and coefficient of variation, should be systematically analysed and compared with expected levels of these indicators. In case of significant deviations, the adoption of corrective actions, such as follow-ups of nonrespondent units and integration with data from administrative sources, should be evaluated. Finally, control surveys or *ad hoc* measurements should be regularly carried out to assess the various components of non-sampling errors (e.g., nonresponse errors and interviewer effect).

A*d hoc* analysis, as well as calculation of quality indicators, are aimed, in the first place, at ensuring the quality of the disseminated estimates and, secondly, at assessing the opportunity to adopt improvement actions for subsequent survey occasions.

[6] For further details see Section II, paragraph 3.2

F. Data storage, dissemination and documentation

F.1. Data storage, dissemination and documentation

Principle F.1. Data storage, dissemination and documentation
Validated microdata should be stored according to standards of the Institute before their dissemination.
The macrodata and microdata disseminated should be pre-emptively treated in order to ensure adequate protection of confidentiality. A dissemination calendar of statistical results should be published.
All phases of the process should be adequately documented.

Guidelines

Before data dissemination, validated microdata should be stored in the repository of the Institute, ARMIDA[7].

Validated microdata should be stored together with the metadata needed for their interpretation (record formats, variables and classifications associated), following the procedure defined by the Institute.

The objective of dissemination is to allow a timely and effective use of the statistical information produced by the Institute, thereby meeting user requirements. To this end, it is useful to define a calendar for the various types of release in advance. Data should be disseminated simultaneously to all users to ensure impartiality and independence of official statistics.

Dissemination of data that are easily accessible and understandable is important in order to allow a better use of them by users. Accessibility is related to the type of media used (on-line dissemination, CD-Rom, paper volume) and to the ease of information retrieval. Given the current national and European directives, the Internet has become the dominant dissemination mode, both through the development of data warehouses and through the publication of online papers, press releases and books. Clarity, however, is linked to the availability of metadata about information content and characteristics of the production process, and quality indicators. Moreover, possible data limitations should be reported, such as the presence of breaks in time series and the possible provisional nature of the data released.
The various types of release, for example, yearbooks and press releases should comply with editorial standards.

The Italian National law establishing the National Statistical System, Legislative decree 322/89, states that respondent confidentiality should be protected and, in particular, that data should be treated appropriately for this purpose before dissemination. In case of dissemination of aggregate data, specific methods may be used, for example the threshold rule, which is set as equal to or greater than three, and methods of perturbation, which consist of perturbing data in order to reduce the possibility of identification and acquisition of information on individual units. With regard to the dissemination of elementary data specific methods may be used, such as recoding of variables to reduce the

[7] The repository ARMIDA (Archive of validated microdata) was established with the main objective of preserving and documenting the data produced by Istat surveys, and has subsequently supported the objective of disseminating data. The data archived in ARMIDA supplies, in fact, the different channels for the dissemination of microdata (for internal use at the Institute through the "Memorandum of access to microdata of ARMIDA" internal users ", for authorities in Sistan, for research files, for standard files, etc.). The micro-data stored in ARMIDA are also used to respond to requests of external users submitted to the Adele laboratory.

information detail, suppression of specific information that may allow the identification of a unit, and methods of perturbation of elementary data. For the protection of confidentiality in data dissemination generalised software should be used.

Documentation of the production process should be produced and archived at all phases of the process, from design to dissemination. Documentation should include process quality indicators, such as indicators of timeliness, coverage and nonresponse, consistency and comparability over time.

Some references

Hundepol A., Domingo-Ferre J., Franconi L., Giessing S., Lenz R., Naylor J., Nordholt E.S., Seri G., De Wolf P.P. (2010) Handbook on Statistical Disclosure Control. Version 1.2. ESSNet SDC – A network of excellence in the European Statistical System in the fields of Statistical Disclosure Control
http://neon.vb.cbs.nl/casc/.%5CSDC_Handbook.pdf

Istat (2004) Metodologie e tecniche di tutela della riservatezza nel rilascio di informazione statistica. Metodi e Norme, N. 20
http://www3.istat.it/dati/catalogo/20040706_00/manuale-tutela_riservatezza.pdf

OMB (2006) Standards and Guidelines for Statistical Surveys. Office for Management and Budget, The White House, Washington, USA.
http://www.whitehouse.gov/sites/default/files/omb/assets/omb/inforeg/statpolicy/standards_stat_surveys.pdf

Section II: Product quality

1. Introduction

Section I describes a series of good practices to be followed in carrying out a statistical process in order to efficiently provide high quality statistics. Nevertheless, planning and carrying out a high-quality process does not exempt from measuring the quality of the statistics produced. Section II lists and comments the criteria used to measure the quality of statistics and to communicate the results of this evaluation to statistics users. No guidelines on how to measure product quality are provided since they have been already introduced in Section I, although in general terms, and because they require a deepening of methodologies which may be found in the specialised literature.

2. The definition and the components of the quality of statistics

In order to measure the quality of statistics, Istat decided to adopt Eurostat definition of quality released in 2003 (ESS Working Group *Assessment of Quality in Statistics*), later resumed in the *European Statistics Code of Practice* (promulgated in 2005 and revised in 2011) and in the Italian Code of Official Statistics (Official Gazette 13 October 2010, n. 240). This definition of quality has become remarkably important since it was included in a legal framework (regulation Ee n. . 223/2009 of the European Parliament and of the Council dated 11 march 2009[8]) on the production of European statistics.

Istat adopted the definition recommended by the LEG on Quality and by Eurostat which states that quality is "the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs" (Eurostat 2002, Eurostat, 2003a). Thus, the quality of the statistics produced and disseminated should be evaluated with respect to the following criteria (Eurostat, 2003a; 2003b)

- relevance
- accuracy
- timeliness and punctuality
- accessibility and clarity
- comparability
- coherence.

It is worth noting that in the Codes such criteria may be joined and ordered in different manners (see Appendices for details), nevertheless without altering their intrinsic meaning. Eurostat definitions of each of these evaluation criteria are listed in Appendix I[9];

[8] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:087:0164:0173:EN:PDF

[9] In these *Guidelines* it was decided to refer to quality criteria names and official definitions as reported by Eurostat, even if ordering and grouping are those used by the European Statistics Code of Practice.

3. Some critical points in measuring the quality of statistics

Measuring the quality of statistics according to the above mentioned components is not an easy task. In fact, it is possible to obtain direct measures (i.e. numeric values) only for a limited number of components (typically timeliness, accuracy and comparability). For the other components, only subjective opinions may be given.

Difficulties arise even for the components which may be measured in quantitative terms. This applies, in particular, to accuracy which consists of different subcomponents whose quantitative evaluation poses methodological and practical problems. The following paragraph describes the error types that may affect accuracy and discusses the concept of reliability, used to measure the quality of statistics that are subject to revision policy[10]. A short paragraph on the type of measurements (or quality indicators) that may be used to measure the effects of sampling and nonsampling errors concludes this section.

3.1 Measuring accuracy and reliability

3.1.1 Accuracy

The level of accuracy is related to the amount of errors that may occur in the production process of survey estimates; the greater the number of errors, the lower the accuracy. In complex sample surveys it is common to distinguish sampling errors (intentional errors) and non-sampling errors (usually not intentional). Non-sampling errors may be: coverage errors, unit and partial nonresponse errors, measurement errors.

All errors may determine an increase of variability associated with final estimates (lower precision) and, in particular, non-sampling errors may be a source of bias (systematic deviations with respect to the true value).

Sampling errors

Sampling errors derive from the choice of carrying the survey exclusively on a sample of the target population. The extent of sampling errors depends mainly on the sample size, the sampling design and the estimator that is going to be used to derive estimates of the quantities of interest. In general, sampling errors tend to decrease by increasing sample size.

Coverage errors

Coverage errors are errors in the lists (sampling frames) used to design and select the sample (in sample surveys) and to contact the units to be observed. Rather frequent coverage errors are: (1) *undercoverage errors*, that is, units belonging to the target population but not listed in the sampling frame, (2) *over coverage errors*, units included in the sampling frame that do not belong to the target population (units not eligible for the survey), (3) *duplications* (units listed several times in the sampling frame). In addition there are errors in the units' identifying variables that may determine unit nonresponse.

[10] The concept of reliability is listed in the European Statistics Code of Practice, while it is not explicitly mentioned in Eurostat quality definition.

Undercoverage errors represent a source of potential bias. Conversely, over coverage, if substantial, may determine an increase in the variability of estimates due to the reduction of the sample size compared to the planned sample (in practice the units selected in the sample that do not belong to the target population are discarded).

Nonresponse errors

Nonresponse errors are due to failed observation of the variables of interest; partial nonresponse occurs when, for the surveyed unit, it is only possible to collect some of the variables of interest; unit nonresponse occurs when no data are collected for the surveyed unit.
Nonresponse errors may cause an increase of variability associated to final estimate and bias.

Measurement errors

They are observation errors that may occur during data collection (measurement errors in strict sense) or the subsequent data treatment phase (revision, registration, coding, monitoring, processing, etc.; also known as data processing errors or processing errors). As a consequence, the available value for a given variable does not correspond to the true value at the end of the survey process. These errors may be a source of bias and may cause a significant increase in variability associated with final survey estimates.

Total survey error

Assessing the impact of errors on final survey estimates is quite difficult. An estimation of bias is possible only in presence of external information or through control surveys. For this reason it is attempted, whenever possible, to prevent bias through targeted actions. When prevention is not effective, it is necessary to adopt methods to reduce bias within acceptable limits.
To estimate variance is less difficult, even though not always possible; typically the part of variability due to sampling error is estimated while variance due to measurement errors is rarely investigated.

It is worth noting that, up to now, despite the advances and research in official statistics, an even approximate estimate of total survey error taking into account all possible survey errors is still far away to be achieved. Indeed the different error types are related to each other and it is difficult to manage this complexity through appropriate statistical models. In most cases, it would be satisfactory to provide an estimate of the impact of the most damaging errors during the production process, using at the same time some methods to compensate the impact of other errors.

3.1.2. Reliability

Reliability is a broader concept than accuracy. It is not only related to final estimates but refers to the sources, tools, methods and procedures used. In general, a process is said to be reliable if it provides essentially the same results (or very close results) when repeated in the same conditions.

It is preferable to assess reliability rather than accuracy, when the production of estimates requires the use of complex processing, using input data from different sources (administrative sources, sample surveys, etc.) that are available at different times. In such cases, it is a common practice to provide preliminary estimates only based on the available data at that time and then update them when new input data become available. The revision policy sets how and when estimates are updated.

Revisions should be accompanied by appropriate indicators to assess their impact on estimates using revision triangles (for details see Istat, 2010a).

3.2. Quality indicators

The difficulties in measuring accuracy or reliability and, more generally, the single components of quality often result in a compromise: few direct measures are usually accompanied by a set of indirect measures. These indirect measures often coincide with some indicators related to the statistical production process and, unlike direct measurements, they do not provide an estimate of bias, or of the increase in variability caused by some error sources. Nevertheless, indirect measurements are relatively easy to obtain and may operate as alarm bells, in addition to helping in monitoring the production process. For these reasons, indirect measurements are widely used in quality management by National Statistical Institutes.
Generally speaking, direct or indirect measurements of the quality components are defined *quality indicators*. The set of quality indicators adopted by Istat is fairly broad[11] and incorporates much of the work done at European level[12].

Some references

Eurostat (2002) Quality in the European Statistical System – The Way Forward. 2002 Edition (Leg on Quality). Luxembourg.
http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/ESS_QUALITY_RECOMMENDATIONS_2002_EN_0_1.pdf

Eurostat (2003a) "Definition of quality in statistics". Working group "Assessment of quality in statistics", Luxembourg, 2-3 October 2003.
http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/ess%20quality%20definition.pdf

Eurostat (2003b) "Standard Quality Report". Methodological Documents, Working Group "Assessment of quality in statistics", Luxembourg, 2-3 October 2003
http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/STANDARD_QUALITY_REPORT_0.pdf

Eurostat (2005) "European Statistics Code of Practice for the National and Community Statistical Authorities". Adopted by the Statistical Programme Committee on 24 February 2005 and promulgated in the Commission recommendation of 25 May 2005
http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/code_practice.pdf

Eurostat (2009) "ESS Handbook for Quality Reports. 2009 Edition". Eurostat: Metholodogies and working papers
http://epp.eurostat.ec.europa.eu/portal/page/portal/lang-en/ver-1/quality/documents/ESQR_FINAL.pdf

[11] The list of the quality indicators used at Istat is managed through the Informative System called SIDI
[12] An updated list is in Section III.5 of Eurostat manual "ESS Handbook for Quality Reports. 2009 Edition". The list of the quality indicators included in the "EURO-SDMX metadata structure" can be found here: http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/Annexes/ESMS_Structure.xls

IMF (2003) Data Quality Assessment Framework — Generic Framework (July 2003 Framework) http://dsbb.imf.org/vgn/images/pdfs/dqrs_genframework.pdf

Istat (2010a) L'analisi delle revisioni delle informazioni statistiche congiunturali. *Approfondimenti*. http://www3.istat.it/salastampa/comunicati/non_calendario/20101013_00/approfondimenti_analisi_revisioni.pdf

Istat (2010b) "Adozione del Codice italiano delle statistiche ufficiali". Comitato di indirizzo e coordinamento dell'informazione statistica - Comstat, Direttiva n. 10, Gazz. Uff. 13 ottobre 2010, n.240. http://www.sistan.it/norme/dir10.html

Appendices

Appendix I

Eurostat's definitions of the quality components[13]

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.

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

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.

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 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.

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.
Accessibility refers to the physical conditions in which users can obtain the data: where to go, how to order, delivery time, pricing policy, marketing conditions (copyright, etc.), availability of micro or macro data, various formats (paper, files, CD-ROM, Internet, …) etc.
Clarity refers to the data's information environment whether the data are accompanied with appropriate metadata, illustrations such as graphs and maps, whether information on their quality are available (including any limitations on use) and the extent to which additional assistance is provided by the producer.

[13] Taken from: Eurostat (2009) "ESS Handbook for Quality Reports. 2009 Edition". Eurostat: Methodologies and working papers,http://epp.eurostat.ec.europa.eu/portal/page/portal/lang-en/ver-1/quality/documents/ESQR_FINAL.pdf

Appendix II

Principles of the European Statistics Code of Practice

Institutional Environment

Principle 1 – Professional independency - Professional independence of statistical authorities from other policy, regulatory or administrative departments and bodies, as well as from private sector operators, ensures the credibility of European Statistics.

Principle 2 - Mandate for data collection - Statistical authorities have a clear legal mandate to collect information for European statistical purposes. Administrations, enterprises and households, and the public at large may be compelled by law to allow access to or deliver data for European statistical purposes at the request of statistical authorities.

Principle 3 - Adequacy of resources - The resources available to statistical authorities are sufficient to meet European Statistics requirements.

Principle 4 – Commitment to quality - Statistical authorities are committed to quality. They systematically and regularly identify strengths and weaknesses to continuously improve process and product quality.

Principle 5 – Statistical confidentiality – 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 are absolutely guaranteed.

Principle 6 - Impartiality and objectivity - Statistical authorities develop, produce and disseminate European Statistics respecting scientific independence and in an objective, professional and transparent manner in which all users are treated equitably.

Statistical Processes

Principle 7 – Sound methodology - Sound methodology underpins quality statistics. This requires adequate tools, procedures and expertise.

Principle 8 - Appropriate statistical procedures – Appropriate statistical procedures, implemented from data collection to data validation, underpin quality statistics.

Principle 9 – Non-excessive burden on respondents - The reporting burden is proportionate to the needs of the users and is not excessive for respondents. The statistical
authorities monitor the response burden and set targets for its reduction over time.

Principle 10 – Cost effectiveness - Resources are used effectively.

Statistical output

Principle 11 - Relevance - European Statistics meet the needs of users.

Principle 12 - Accuracy and reliability - European Statistics accurately and reliably portray reality.

Principle 13 - Timeliness and punctuality - European Statistics are released in a timely and punctual manner.

Principle 14 - Coherence and comparability - European Statistics are consistent internally, over time and comparable between regions and countries; it is possible to combine and make joint use of related data from different sources.

Principle 15 - Accessibility and clarity – European Statistics are presented in a clear and understandable form, released in a suitable and convenient manner, available and accessible on an impartial basis with supporting metadata and guidance.

Appendix III

Italian Code of Official Statistics
(Directive n. 10/COMSTAT)
Published in the Italian *Gazzetta Ufficiale* 13 October 2010, n. 240

The Italian Code of Official Statistics is based on the principle of the European Code and is statutory for the National Statistical System. It has been adopted in 2010.

The Italian Code of Official Statistics is available at (Italian version only): http://www.sistan.it/codice_statistica.pdf

Appendix IV

Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009

Article 12

Statistical quality

1. To guarantee the quality of results, European statistics shall be developed, produced and disseminated on the basis of uniform standards and of harmonized methods. In this respect, the following quality criteria shall apply:

(a) 'relevance', which refers to the degree to which statistics meet current and potential needs of the users;
(b) 'accuracy', which refers to the closeness of estimates to the unknown true values;
(c) 'timeliness', which refers to the period between the availability of the information and the event or phenomenon it describes;
(d) 'punctuality', which refers to the delay between the date of the release of the data and the target date (the date by which the data should have been delivered);
(e) 'accessibility' and 'clarity', which refer to the conditions and modalities by which users can obtain, use and interpret data;
(f) 'comparability', which refers to the measurement of the impact of differences in applied statistical concepts, measurement tools and procedures where statistics are compared between geographical areas, sectoral domains or over time;
(g) 'coherence', which refers to the adequacy of the data to be reliably combined in different ways and for various uses.

2. In applying the quality criteria laid down in paragraph 1 of this Article to the data covered by sectoral legislation in specific statistical domains, the modalities, structure and periodicity of quality reports provided for in sectoral legislation shall be defined by the Commission in accordance with the regulatory procedure referred to in Article 27(2).

Specific quality requirements, such as target values and minimum standards for the statistical production, may be laid down in sectoral legislation. Where sectoral legislation does not so provide, measures may be adopted by the Commission. Those measures, designed to amend non-essential elements of this Regulation by supplementing it, shall be adopted in accordance with the regulatory procedure with scrutiny referred to in Article 27(3).

3. Member States shall provide the Commission (Eurostat) with reports on the quality of the data transmitted. The Commission (Eurostat) shall assess the quality of data transmitted and shall prepare and publish reports on the quality of European statistics.