

Catalogue no. 12-539-X

Statistics Canada Quality Guidelines

Fifth Edition – October 2009



How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website at www.statcan.gc.ca, e-mail us at infostats@statcan.gc.ca, or telephone us, Monday to Friday from 8:30 a.m. to 4:30 p.m., at the following numbers.

Statistics Canada's National Contact Centre

Toll-free telephone (Canada and United States):

Inquiries line	1-800-263-1136
National telecommunications device for the hearing impaired	1-800-363-7629
Fax line	1-877-287-4369

Local or international calls:

Inquiries line	1-613-951-8116
Fax line	1-613-951-0581

Depository Services Program

Inquiries line	1-800-635-7943
Fax line	1-800-565-7757

Information to access the product

This product, catalogue no. 12-539-X, is available free in electronic format. To obtain a single issue, visit our website at www.statcan.gc.ca and select "Publications."

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "About us" > "Providing services to Canadians."

Statistics Canada Quality Guidelines

Fifth Edition – October 2009

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2009

All rights reserved. The content of this electronic publication may be reproduced, in whole or in part, and by any means, without further permission from Statistics Canada, subject to the following conditions: that it be done solely for the purposes of private study, research, criticism, review or newspaper summary, and/or for non-commercial purposes; and that Statistics Canada be fully acknowledged as follows: Source (or "Adapted from", if appropriate): Statistics Canada, year of publication, name of product, catalogue number, volume and issue numbers, reference period and page(s). Otherwise, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form, by any means—electronic, mechanical or photocopy—or for any purposes without prior written permission of Licensing Services, Client Services Division, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6.

December 2009

Catalogue no. 12-539-X
ISSN 1708-6256

Frequency: Occasional

Ottawa

Cette publication est disponible en français.

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Preface

Canadians and their government have established Statistics Canada to have access to a trusted source of information. Trust can only be established if the data Statistics Canada produces are consistent with the country's needs and representative of the world we are seeking to describe. In other words, information has to be relevant and of high quality.

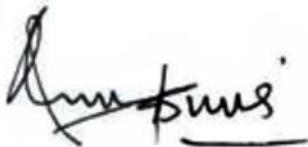
Quality is, therefore, fundamental to Statistics Canada's mandate to produce information. There is an essential characteristic of quality that needs to be understood: quality, defined as representativeness of the universe we are trying to capture, will deteriorate automatically in the absence of pro-active action. This is because, as the world around us evolves, our methods to maintain representativeness of our data must evolve as well.

In view of these realities, Statistics Canada has a long tradition of providing guidance in its survey designs by consolidating its experiences and conclusions about what constitute "best practices" into a set of Quality Guidelines. The first edition of Quality Guidelines appeared in 1985. Revised editions were released in 1987, 1998 and 2003. In keeping with the need to keep the guidelines evergreen, the present document has been significantly updated from the previous edition to reflect further advances in survey methodology over the past six years.

The guidelines presented in this document do not all apply equally to every data acquisition process. Their applicability and importance must be carefully considered in light of the particular requirements and constraints of individual programs. This document must, therefore, be used with professional care and judgment.

While the guidelines provided in this document are no substitute for expertise and judgment on the part of survey design staff, the underlying concern for quality must pervade all our activities. All staff involved in statistical activities are responsible for ensuring that quality has high priority in the design and implementation of statistical methods and procedures under their control.

I want to thank many Statistics Canada experts who have contributed to the preparation of the Quality Guidelines over many years. The guidance of the Methods and Standards Committee helped to make this a better document.



Munir A. Sheikh
Chief Statistician

Table of contents

Page	
Introduction	6
Figure 1 Levels 1 and 2 of the Generic Statistical Business Process Model	9
Survey steps	13
1. Objectives, uses and users	14
2. Concepts, variables and classifications	16
3. Coverage and frames	19
4. Sample design	23
5. Questionnaire design	28
6. Data capture, collection and coding	32
7. Use of administrative data	40
8. Response and non-response	46
9. Editing	51
10. Imputation	54
11. Weighting and estimation	59
12. Seasonal adjustment and trend-cycle estimation	63
13. Benchmarking and related techniques	68
14. Data quality evaluation	71
15. Disclosure control	75
16. Data dissemination and communication	79
17. Data analysis and presentation	82
18. Documentation	87

Introduction

Statistical information is critical to the functioning of a modern democracy. Without good data, the quality of decision-making, the allocation of billions of dollars of resources, and the ability of governments, businesses, institutions and the general public to understand the social and economic reality of the country would be severely impaired. A national statistical agency, such as Statistics Canada, plays an essential role in the production and dissemination of statistical information.

The credibility of a statistical agency in fulfilling this key role rests on the following pillars: production of high quality statistical information; cost efficiency; privacy; confidentiality; and the maintenance of a highly capable and motivated workforce.

More specifically, the quality of the information it produces, its relevance in particular, is of fundamental importance to a statistical agency. Unless the statistical agency is capable of producing high quality data, both the users and the suppliers of statistical data would soon lose confidence in the statistical agency, making its job impossible. As an introduction to the guidelines, this section presents the principles of quality assurance at Statistics Canada within which they are applied.

Principles of quality assurance at Statistics Canada

The Agency's management structure, policies and guidelines, consultative mechanisms, project development and management approach, and environment have been developed to facilitate and assure effective management of quality. The basic mechanisms for managing quality are described in Statistics Canada's Quality Assurance Framework (Statistics Canada, 2002c).

The framework consists of a wide variety of mechanisms and processes acting at various levels throughout the Agency's programs and across its organization. The effectiveness of this framework depends not on any one mechanism or process but on the collective effect of many interdependent measures. These build on the professional interests and motivation of the staff. They reinforce each other as means to serve client needs. They emphasize the Agency's objective professionalism, and reflect a concern for data quality. An important feature of this strategy is the synergy resulting from the many players in the Agency's programs operating within a framework of coherent processes and consistent messages. Within the framework, the Quality Guidelines provide an accompanying document that describes a set of best practices for all of the "steps" of a statistical program, and is aimed at project team members who are charged with the development and implementation of statistical programs.

Underlying all of these mechanisms, processes and practices are eight guiding principles.

Quality is relative, not absolute

A significant feature of the management of quality is the balancing of quality objectives against the constraints of financial and human resources, the goodwill of respondents in providing source data, and competing demands for greater quantities and detail of information. The management of quality is not the maximization of quality at all costs, but the achievement of an appropriate balance between the quantity and quality of information yielded by the Agency's programs and the resources available. Within individual programs the challenge is to make the appropriate trade-offs between the evolving needs of clients, costs, respondent burden, and the various elements or dimensions of quality.

Statistical data are important because of the use to which they are put. It follows that the quality of statistical data can only be judged against their relevance and how well they represent the world we seek to describe. It also follows that it is important for the statistical agency to have a thorough understanding of the uses to which its data are put, and to do so it must maintain ongoing relations with its user community.

This principle also recognizes that achieving “perfect” quality is neither desirable nor affordable (in fact it is rarely even possible). Data are subject to numerous sources of error, both sampling and non-sampling, and it is the job of the statistical agency to balance factors such as accuracy, cost and burden on respondents in developing a statistical program. Minimizing error itself is not the goal; each statistical program must be designed within the context of what is feasible and how important the data are to users.

Statistics Canada strives to build relevance and quality into all its programs and products. The quality of its official statistics is founded on the use of sound scientific methods adapted over time to changing client needs, to the changing reality that the Agency aims to measure, and to the capacity or willingness of respondents to supply reliable and timely data. The Quality Guidelines are one of the tools that will aid in building quality into the design of each program.

Quality is multidimensional

During the past twenty years, statistical agencies have arrived at a consensus that the concept of “quality” of statistical information is multi-dimensional. Statistics Canada defines quality in terms of six dimensions; other statistical agencies and organizations have defined similar frameworks. While these definitions may differ slightly, they all recognize that there is no one single measure of data quality.

At Statistics Canada, the dimensions of quality are defined as follows:

The relevance of statistical information reflects the degree to which it meets the real needs of clients. It is concerned with whether the available information sheds light on the issues of most importance to users. Hence, relevance is the most important dimension of quality; one could even consider it among the pillars of a statistical agency. It is largely in the domain of users of the information; it is not something that a statistical agency can establish by itself. Comparatively, the other dimensions of quality are much more within the control of the statistical agency.

The accuracy of statistical information is the degree to which the information correctly describes the phenomena it was designed to measure. It is usually characterized in terms of error in statistical estimates and is traditionally decomposed into bias (systematic error) and variance (random error) components. It may also be described in terms of the major sources of error that potentially cause inaccuracy (e.g., coverage, sampling, non-response, response).

The timeliness of statistical information refers to the delay between the reference point (or the end of the reference period) to which the information pertains, and the date on which the information becomes available. It is typically involved in a trade-off against accuracy. The timeliness of information will influence its relevance.

The accessibility of statistical information refers to the ease with which it can be obtained from the Agency. This includes the ease with which the existence of information can be ascertained, as well as the suitability of the form or medium through which the information can be accessed. The cost of the information may also be an aspect of accessibility for some users.

The interpretability of statistical information reflects the availability of the supplementary information and metadata necessary to interpret and utilize it appropriately. This information normally covers the underlying concepts, variables and classifications used, the methodology of data collection and processing, and indications of the accuracy of the statistical information.

The coherence of statistical information reflects the degree to which it can be successfully brought together with other statistical information within a broad analytic framework and over time. The use of standard concepts, classifications and target populations promotes coherence, as does the use of common methodology across surveys. Coherence does not necessarily imply full numerical consistency.

These dimensions are overlapping and interrelated; in managing quality all of them must be considered. A failure in any one dimension will result in the failure of the entire statistical program.

Every employee has a role to play in assuring quality

Through its policies, guidelines and internal communications, Statistics Canada has made it clear to its employees that everyone has a role to play in assuring quality, from the employees working on daily production tasks to the highest level of management. This reflects the philosophy of Deming (1982) that quality is not something that can be “inspected into” the process, but should be built into the process to begin with. Thus there is no one body at Statistics Canada explicitly charged with quality assurance.

As well, the human resources practices of the agency reflect the principle that quality is everyone’s business. The recruitment, training and development programs of the Agency put considerable emphasis on technical competencies and an understanding of what constitutes high quality data.

Balancing of the dimensions of quality is best achieved through a project team approach

Because quality is multidimensional, it follows that different dimensions tend to be the area of expertise of different groups in the Agency. For many years, Statistics Canada has realized that the tradeoffs among the various dimensions of quality are best achieved through a project team approach.

The management of quality at Statistics Canada occurs within a matrix management framework – project management operating within the functional organization. The Agency is functionally organized into six Fields. Three of these are primarily responsible for statistical programs of data production and analysis in various subject-matter areas (e.g. social statistics, business statistics and national accounts). The other three Fields are primarily involved in the provision of infrastructure and services to be used by the statistical programs (e.g. methodology, informatics, collection operations, dissemination and management systems). A typical statistical program is managed by one of the subject-matter divisions and draws heavily on the resources of infrastructure and service areas for inputs to the program.

The use of an interdisciplinary project team approach for the design or redesign of a statistical program is important in ensuring that quality considerations relating to all the components and steps in the program receive appropriate attention during design, implementation and assessment. It is the responsibility of the functional organizations to ensure that project teams are adequately staffed with people able to speak with expertise and authority for their functional area. Subject-matter staff bring knowledge of content, client needs, and relevance. Methodologists bring their expertise in statistical methods and data quality trade-offs, especially with respect to accuracy, timeliness and cost. Operations experts bring experience in operational methods, and concerns for practicality, efficiency, field staff and respondents. The system experts bring a systems view, and knowledge of technology standards and tools.

It is within such a project team that the many decisions and trade-offs necessary to ensure an appropriate balance between concern for quality and considerations of cost and response burden are made. Together the team has to balance the conflicting pressures in order to develop an optimal design. The fact that each member of the team is a part of a specialized functional organization, from which a variety of more specialized and management resources can be called upon when warranted, helps in resolving both technical challenges and conflicts arising in a project.

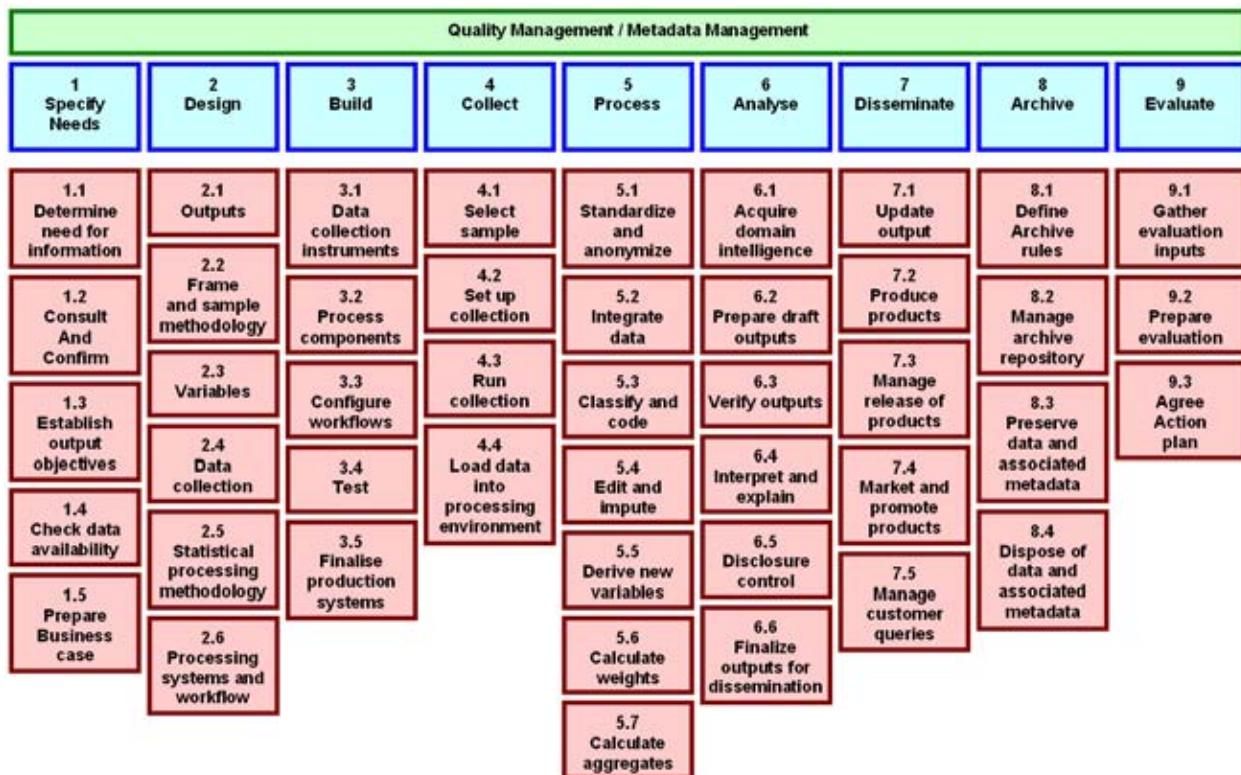
Projects are normally guided by a more senior Steering Committee that may include managers from each of the major participating areas. This Committee, which is part of the formal approval mechanism for the design and implementation of the program, provides overall guidance, broad budgetary and design parameters, and helps to ensure that appropriate resources are available to the project. It also provides a forum for resolving any issues that cannot be satisfactorily resolved within the project team.

Quality must be built in at each phase of the process

Statistical agencies have often modeled the statistical process, as an aid in managing it. The second edition (April 1987) of Statistics Canada's Quality Guidelines contained a schematic of the statistical survey process. More recently, agencies such as Statistics New Zealand, the Australian Bureau of Statistics, Statistics Sweden, Statistics Norway, Statistics Netherlands, and the Joint UNECE / Eurostat / OECD Work Sessions on Statistical Metadata (METIS) have developed various drafts of a Generic Statistical Business Process Model (GSBPM). The model developed by METIS (UNECE Secretariat 2008), which is based on that of Statistics New Zealand but with input from several other agencies, including Statistics Canada, is shown in Figure 1.

These various models have in common the division of the process into a number of phases or steps. While the details of the various models vary, all contain common elements: the specification of user needs, the design of the program, the implementation or "build" phase (specifications, systems, operations manuals, training, etc.), the execution phase (collection, verification, etc.) and the evaluation phase. A basic principle of quality assurance is that it must be considered at all of these phases. If the user needs are not understood or are incorrectly specified, then all the steps that follow will only result in data that are not relevant. Measures must also be taken to ensure that the design is done properly; if it is not then no amount of perfect implementation and execution will compensate. However good design is not enough; if it is not implemented or executed correctly then the good design has gone to waste. And without proper evaluation the statistical agency will not know whether the statistical program has met its objectives or not.

Figure 1 Levels 1 and 2 of the Generic Statistical Business Process Model



The principle that quality must be built into each phase, together with the notion that quality is multidimensional, leads logically to the conceptualization of quality assurance management as a matrix defined by the dimensions of quality (relevance, accuracy, etc.) in one dimension and the phases of a survey (specification of needs, design, implementation, execution, evaluation) in the other dimension. A comprehensive approach to the management of quality demands that all the cells of this matrix be considered.

Quality assurance measures must be adapted to the specific program

At Statistics Canada, statistical program managers are responsible and accountable for delivering their programs. Each statistical activity manager has the responsibility to ensure that the Agency's concern for quality is adequately reflected in the statistical program's methods and procedures. It has always been clear that the Quality Guidelines are exactly that: guidelines to assist program managers, and not rules to be followed. It is clearly not the expectation that every program will adhere to every guideline; such a proposition would be prohibitively expensive and unnecessary, given the variation in the importance of the various statistical programs. Instead, statistical program managers and the project teams that support them are expected to make the necessary decisions.

Users must be informed of data quality so that they can judge whether the statistical information is appropriate for their particular use

Finally, in order for the user to make informed use of the statistical information provided, he or she must be able to assess whether the data are of sufficient quality. For some dimensions of quality, such as timeliness, users are able to assess the quality for themselves. Other aspects, such as interpretability, coherence and even relevance may not be as obvious. The dimension of accuracy in particular is one which users may often have no way of assessing and must rely on the statistical agency for guidance.

Over the years the Agency has developed policies and tools to assist users. It has long had a Policy on Informing Users of Data Quality and Methodology (PIUDQM), which prescribes minimum amounts of information on data quality and methodology that are to be provided to users. All data are released through The Daily and are accompanied by a link to the Integrated Meta Database (IMDB) that provides information on concepts, definitions, data sources and methodology for each statistical program.

Concerning accuracy, the Policy also specifies that all data releases are to be accompanied by information on three common sources of error: coverage (the difference between the target population and the frame used to conduct the survey), non-response (the portion of the sample who did not respond), and sampling error (a source of error when a sample rather than a census was conducted), as well as any other significant sources of error (e.g., response errors, processing errors, errors introduced for disclosure control).

Quality must be at the forefront of all activities

Finally, unless proactive actions are taken, the quality of data should, and does, deteriorate over time. For example, a "relevance gap" often occurs given the time it takes between the emergence of a need for data and the ability to produce it. A gap on other quality dimensions can also manifest itself as response rates decline over time due to changes in societal attitudes, or as systems become outdated or methodology becomes in need of redesign.

A statistical agency must keep quality at the forefront of all its activities in order to minimize the relevance gap and prevent a significant decrease in quality over time.

The quality of program processes and outputs must be assessed, checked or reviewed constantly. Such review mechanisms must be integrated into the agency's planning and decision making processes. Statistics Canada's National Statistics Council and senior bilateral arrangements with key departments and agencies (effective dialogue with stakeholders), Integrated Program Reporting (systematic assessment of statistical outputs) and Quality Review Program (formal assessment of statistical processes) are examples of such quality assessment or review mechanisms.

Purpose and scope of the guidelines

Section 2 of this document brings together guidelines and checklists on many issues that need to be considered in the pursuit of quality objectives in the execution of statistical activities. Its focus is on how to assure quality through effective and appropriate design and implementation of a statistical program from inception through to data evaluation, documentation and dissemination. These guidelines draw on the collective knowledge and experience of many Statistics Canada employees. It is expected that guidelines will be useful to staff engaged in the planning and design of surveys and other statistical programs, as well as to those who evaluate, analyze and use the outputs of these programs.

The main purpose of the Quality Guidelines is to provide a comprehensive list of guiding principles and good practices in survey design. To better appreciate the scope of the guidelines, it is important to define its use of the words survey and design.

The term survey is used generically to cover any activity that collects or acquires statistical data. Included are:

- a census, which attempts to collect data from all members of a population;
- a sample survey, in which data are collected from a (usually random) sample of population members;
- collection of data from administrative records, in which data are derived from records originally kept for non-statistical purposes;
- a derived statistical activity, in which data are estimated, modeled, or otherwise derived from existing statistical data sources.

The guidelines are written with censuses and sample surveys as the main focus. The quality of derived statistical activities is, of course, largely determined by the quality of the component parts, and as such, derived statistical activities are not the direct focus of this document.

The term design is used to cover the delineation of all aspects of a survey from the establishment of a need for data to the production of final outputs (the microdata file, statistical series, and analysis).

The core of this document (Section 2) concentrates on quality issues as they relate to the design of individual surveys. It is, however, important to keep in mind that the context in which each individual survey is developed imposes constraints on its design. Each new survey, while aiming to satisfy some immediate information needs, is also contributing information to a base of statistical data that may be used for a range of purposes that go well beyond those identified at the time of the survey's design. It is therefore important to ensure that the output from each individual survey can, to the extent possible, be integrated with, and used in conjunction with, data on related topics derived from other surveys. This implies a need to consider and respect the statistical standards on content or subject-matter that have been put in place to achieve coherence and harmony of data within the national statistical system. These include statistical frameworks (such as the System of National Accounts), statistical classification systems (such as those for industry or geography), as well as other concepts and definitions that specify the statistical variables to be measured. The usefulness of new statistical data is enhanced to the extent that they can be utilized in conjunction with existing data.

References

- Brackstone, G. 1999. "Managing data quality in a statistical agency." *Survey Methodology*. Vol. 25. p. 139-149.
- Deming, W.E. 1982 *Quality, Productivity, and Competitive Position*, Cambridge, MA: Massachusetts Institute of Technology.
- Fellegi, I. 1996. "Characteristics of an effective statistical system." *International Statistical Review*. Vol. 64. p. 165-197.
- Statistics Canada. 1987 *Quality Guidelines*. Second Edition. Ottawa, Ontario.
- Statistics Canada. 2000c. "Integrated Metadatabase – Guidelines for Authors." *Standards Division Internal Communications Network*. No date. <http://stdsweb/standards/imdb/imdb-menu.htm>.
- Statistics Canada. 2000d. "Policy on Informing Users of Data Quality and Methodology." *Statistics Canada Policy Manual*. Section 2.3. Last updated March 4, 2009. http://icn-rci.statcan.ca/10/10c/10c_010_e.htm.
- Statistics Canada. 2002c. *Statistics Canada's Quality Assurance Framework - 2002*. Statistics Canada Catalogue no. 12-586-X1E. http://method/Branch/Committees/MethodsAndStandards/QualityAssuranceFramework_e.htm.
- Statistics Canada. 2003d. *Statistics Canada Policy Manual*. Last updated May 27, 2009. http://icn-rci.statcan.ca/10/10c/10c_000_e.htm.
- Trewin, D. 2002. "The importance of a quality culture." *Survey Methodology*. Vol. 28. p. 125-133.
- UNECE Secretariat. 2008. "Generic Statistical Business Process Model: Version 3.1 – December 2008". *Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS)*.

Survey steps

This section is organized in subsections that correspond to the main activities of a typical survey. The subsections all follow the same structure, describing the Scope and purpose, Principles, Guidelines and Quality indicators related to each activity, as defined below. The first subsection addresses the stage at which objectives, uses and users are identified. The subsections that follow cover other survey steps roughly in the chronological order in which they would typically take place. However, there are significant interdependencies between some steps such as, for example, between questionnaire design and data collection and capture operations. As well, survey steps as discussed here do not always proceed strictly sequentially. Some activities can take place concurrently, for example, frame development, sampling plans and questionnaire design. Other steps, such as data quality evaluation and documentation touch on most other activities and do not take place as discrete activities on their own.

Scope and purpose

Under the heading of Scope and purpose, a description of the concepts and key terms of the main activity or survey step are provided. The objective of the step and why it is important are stated briefly.

Principles

Principles are the broad, underlying policies, directions and approaches which govern the design of the activity in question, with emphasis on those that relate to quality.

Guidelines

Guidelines are known good practices that have evolved in the design and implementation of statistical surveys. Not all of these guidelines can be applied to every survey. They provide checklists to aid survey design. Judgment is still needed in deciding how to weigh the considerations that these Guidelines suggest.

On the other hand, Statistics Canada does have policies that have a bearing on many aspects of statistical activities in the Agency, and which may place requirements on the way particular activities are carried out. These are documented separately in the Statistics Canada Policy Manual. Wherever a policy has a bearing on a particular topic covered by these Guidelines, the existence and relevance of the policy is indicated.

Quality indicators

Quality measures give a direct measure of the quality of data but, in practice, they can rarely be explicitly calculated. For example, in the case of accuracy, it is almost impossible to measure non-response bias, as the characteristics of those who do not respond can be difficult to ascertain. Instead, certain information can be provided to help “indicate” quality. Quality indicators usually consist of information that is a by-product of the statistical process. They do not measure quality directly but can provide enough information to offer valuable insight into quality. Included in this section are both quality measures, when they exist, and quality indicators.

Information presented in this section will be useful to methodologists tasked with producing quality measures to accompany the statistical outputs. It will also be of interest to director of program areas, survey managers and data users, who will use the indicators to assess and compare the quality of statistical products. In addition, this section will be of interest to the directors of program areas and production managers, as it will provide a basis for monitoring performance in terms of the quality of the processes and products in the program area.

1 Objectives, uses and users

1.1 Scope and purpose

Objectives are the purposes for which information is required, stated within the context of the program, research problem or hypotheses that gave rise to the need for information. Uses narrow down and specify more precisely the information needs, for example, by describing what decisions may be made based on the information collected and how such information will support these decisions. Users are the organizations, agencies, groups or individuals expected to use the project deliverables.

The first task in planning a statistical activity is to specify the objectives. A clear statement of objectives guides all subsequent steps and could be revised many times during the survey development. Meeting the objectives might require mounting a new survey, redesigning an existing survey, using existing data products, using administrative records, or a combination of these activities. The information needs as stated in the objectives must justify any response burden that will be generated. The relevance of project deliverables to the targeted user community must also be clearly stated.

1.2 Principles

The hypotheses to be tested, specific data requirements and use of the data, data quality expectations, budget constraints and expected delivery dates should all be stated in the objectives. The concept, definition, unit of analysis and the target population, which are discussed in the subsequent sections, should also be defined at this stage. This will allow the intended as well as other potential data users to determine if and to what extent the project deliverables meet their needs.

1.3 Guidelines

1.3.1 Planning

- Develop survey objectives and constraints in partnership with important users and stakeholders. Establish and maintain relationships with users of information in the private and public sectors and with the general public to enhance the relevance of the information produced and to improve the marketing of products and services. Among important users are representatives of potential markets, policy makers and agents who require the information for legislated use. Before major designs or redesigns, consider conducting a feasibility study and/or pilot test. Conduct extensive and user-focused consultation routinely so as to identify content options, survey relevancy, to determine the need for a cross-sectional or a longitudinal survey as well as to develop public support for the program when it reaches the data collection stage.
- Focus analysis of user needs and data requirements on finding the most cost-effective solutions for both the short and long term in the context of statistical framework. Having the users specify their analytical plan or proposed release tables in advance will help to more clearly define their detailed requirements. Before embarking on the design of a new statistical activity (or redesigning an existing one, longitudinal or cross-sectional survey), analyze currently available statistics in the area in terms of sources, frequency, quality, timeliness, etc. Determine the best trade-off between adequacy of the available statistics to meet the requirements of clients and the cost and time required to undertake a new activity to produce statistics that do not already exist.
- Where explicit data quality targets exist, include them in the statement of survey objectives in terms of measurable aspects of quality for the entire population or specific domains. Targets can be set in terms of measures such as sampling error, coverage rates, response rates, and timeliness. With administrative data and derived statistical activities, the data quality will be directly related to the quality of input data sources.

- During the planning process, clarify the operational constraints such as time frame, costs, resources and data collection methods. Other issues such as the use of proxy responses, respondent recall and the need to measure seasonal data variation should also be taken into account. This is conducted in phases of increasing detail and exactitude. To start with, estimates are based on broad assumptions of the methodology to be used. These estimates have to be more exact and detailed at each stage of planning. They should be based on historic information when available and updated as the objectives are reviewed.
- In determining the extent to which a survey will meet user needs, seek a reasonable trade-off between these needs, the survey objectives and the budget, response burden and privacy considerations. Although the Agency may have little discretion where a legal requirement is in place, in other cases it is worthwhile to formulate alternative methodological approaches, means and modes of data collection, frequencies, geographical detail, etc. with a view to arriving at an optimum solution. It might also be necessary to conduct a pilot survey or feasibility study to find an optimum solution.
- Review ongoing statistical activities at regular intervals. Statistical programs need to evolve, adapt and innovate so as to keep pace with the changing demands of the users they serve or demands of new users. The purpose of the activity or its statement of objectives needs to be reviewed periodically to enhance the relevance of the statistical product to user needs and constraints (budget, time frame, resources, etc.), as they evolve or change. Sometimes the redesign overhaul of existing surveys may be desirable to maintain the reliability of key statistical series, especially if sources of information have changed or the way in which they are made available is reengineered or rethought.

1.4 Quality indicators

Main quality element: relevance

Describe and classify key users of project deliverables.

Describe the needs of key users and their anticipated uses of data products, in terms of analytical plans and release tables. Address any gaps between needs and deliverables.

If changes are made to a survey program where estimates are fitted into a time series, evaluate the impact on the time series of those changes.

References

Blanc, M., Radermacher, W. and Körner, T. 2001. "Quality and users." *International Conference on Quality in Official Statistics 2001*. Session 15.1. Stockholm, Sweden.

Brackstone, G.J. 1993. "Data relevance: keeping pace with user needs." *Journal of Official Statistics*. Vol. 9, no. 1, p. 49-56.

Statistics Canada. 2003. *Survey methods and practices*. Statistics Canada Catalogue no. 12-587 XPE. Ottawa, Ontario. 396 p.

2 Concepts, variables and classifications

2.1 Scope and purpose

Concepts are general or abstract ideas that express the social and/or economic phenomenon to be studied. They are the subjects of inquiry and analysis that are of interest to users.

A variable consists of two components, a statistical unit and a property. A statistical unit is the unit of observation or measurement for which data are collected or derived (e.g. persons or households in social surveys, and enterprises or establishments in business surveys) (Statistics Canada, 2008). A property is a characteristic or attribute of the statistical unit. Definitions of variables must be unambiguous and clearly specified in the context of the analytical purposes for which the data are to be collected (Statistics Canada, 2004).

Classification is a systematic grouping of values that a variable can take comprising mutually exclusive classes, covering the full set of values, often providing a hierarchical structure for aggregating data so as to facilitate analysis and interpretation. More than one classification can be used to represent data for a given variable (Statistics Canada, 2004).

2.2 Principles

Statistics Canada aims to ensure that the information it produces provides a consistent and coherent picture of the Canadian economy, society and environment, and that its various datasets can be analyzed together and in combination with information from other sources. To achieve this the Agency follows conceptual frameworks, makes use of standard names and definitions for populations, statistical units, concepts, variables and classifications in statistical programs and uses consistent collection and processing methods for the production of statistical data across surveys. The Statistics Canada Policy on Standards governs how this is to be done (Statistics Canada, 2004). Where applicable, three types of standards should be respected, in descending order of compulsion; departmental standards, recommended standards and program specific standards (Statistics Canada, 2004).

2.3 Guidelines

2.3.1 Using standards

- Specify concepts and variables clearly and relate them to their intended use. Emphasis should be placed on the use of standard definitions of concepts, variables, classifications, statistical units and populations established under the Statistics Canada Policy on Standards (Statistics Canada, 2004). In choosing naming conventions, take into account the similarity or dissimilarity of existing standards and usage. Use titles from existing standards only for what is defined in the standards.
- Use standard definitions to make it possible to compare data collected from different sources and to integrate data across sources (Statistics Canada, 2004). Statistics Canada has standard classifications of industries, products, instructional programs, occupations, financial accounting and geography (Statistics Canada 2007a (NAICS), 2007b (NAPCS), 2000 (CIP), 2006a (NOC-S), 2006b (COA) and 2007c (SGC) as well as of a large number of other domains used for social and economic statistics.
- In addition to Statistics Canada's standard classifications, there are international standard classifications produced by the United Nations Statistical Office, the International Labour Office, Eurostat, and other international and regional agencies. The Standards Division has produced official concordances to a number of international standard classifications. When there is a requirement to provide data to international agencies, use official concordances when they are available.
- Use standard units of observation to facilitate the comparison of data. Classifications are usually designed with particular units of observation in mind. For example, the North American Industry Classification Systems (NAICS) is designed primarily for classifying establishments.

- Be aware of derived statistical activities or statistical frameworks (e.g., the System of National Accounts) whose definitions of concepts and variables may have a significant effect on specific data collection activities (Statistics Canada, 1989).
- Sometimes, there is more than one way to measure a concept. The variables and classifications chosen to measure a concept will also need to take into account factors such as the ease of obtaining the information required, the respondent burden imposed, the collection method, the context in which the question(s) must be asked, the processing of the data (especially editing, imputation and weighting techniques), whether the information can be obtained from administrative records, and the costs associated with collection and processing. Thus, the measurement approach adopted may be more or less successful in providing the desired interpretation of the concept. A variable chosen at one point in time may become obsolete later if new factors come into play and may therefore need to be modified or changed. Therefore, it is important to ensure that the latest approved version of the variable is used. Updated standards are made available on the Statistics Canada website.
- In the absence of an official standard, examine the concepts, variables and classifications being used by related statistical programs and consult with the Standards Division when necessary.

2.3.2 Using classifications

- To maximize flexibility of use, code microdata and maintain files at the lowest possible level of the appropriate classification. Aggregation at a higher level may be required for particular analytical purposes or to satisfy confidentiality or data reliability constraints. Wherever possible, use the hierarchy of the classification in terms of the classes or higher-level aggregations of the standard. If this is not possible follow a common collapsing strategy for aggregation and document differences between the standard and adopted levels of classifications/aggregations used. Use classifications that reflect both the most detailed and the collapsed levels. Make clear to users how these fit into higher-level (e.g. less detailed) classifications.

2.4 Quality indicators

Main quality elements: coherence, interpretability, relevance

Describe key statistical concepts, including the statistical measure, the population, variables, units, domains and time reference. This information gives users an understanding of the relevance of the output to their needs.

Provide accurate references when standard concepts, variables and classifications are adopted.

Describe, justify and if possible, measure (qualitatively if not quantitatively) any departures from standards. This gives users an indicator of relevance, and aids interpretability.

References

Statistics Canada. 1989. *A User Guide to the Canadian System of National Accounts*. Statistics Canada Catalogue no. 13-589. Ottawa, Ontario. 106 p.

Statistics Canada. 2000. *Classification of Instructional Programs (CIP)*. Last updated December 21, 2007. <http://www.statcan.gc.ca/concepts/classification-eng.htm>.

Statistics Canada. 2004. "Policy on Standards." *Statistics Canada Policy Manual*. Section 2.10. Last updated March 4, 2009. http://icn-rci.statcan.ca/10/10c/10c_014_e.htm.

Statistics Canada. 2006a. *National Occupational Classification for Statistics (NOC-S)*. Statistics Canada Catalogue no. 12-583-XIE. Ottawa, Ontario. 648 p.

Statistics Canada. 2006b. *Chart of Accounts: Standard for Reporting on the Financial Position and Performance of a Business*. Last updated July 3, 2006. <http://stdsweb/english/Subjects/Standard/coa-standdraft.htm>.

Statistics Canada. 2007a. *North American Industry Classification System - NAICS Canada, 2007*. Last updated March 10, 2009. <http://www.statcan.gc.ca/subjects-sujets/standard-norme/naics-scian/2007/index-indexe-eng.htm>.

Statistics Canada. 2007b. *North American Product Classification System (NAPCS), 2007*. Last updated August 15, 2008. <http://www.statcan.gc.ca/subjects-sujets/standard-norme/napcs-scan/napcs-scan-eng.htm>.

Statistics Canada. 2007c. *Standard Geographical Classification, SGC 2006*. 2 vols. Last updated August 1, 2008. <http://www.statcan.gc.ca/subjects-sujets/standard-norme/sgc-cgt/geography-geographie-eng.htm>.

Statistics Canada. 2008. *Industry Classification Coding System (ICCS) Version 1.4, (NAICS 2007)*. Statistics Canada Catalogue no. 12F0074XCB. Ottawa, Ontario.

Statistics Canada. 2008. *Standard Statistical Units*. Last updated July 25, 2008. <http://www.statcan.ca/english/concepts/stat-unit-def.htm>.

Statistics Canada. No date. *Integrated Metadatabase (IMDB) - Guidelines for Authors*. No date. <http://stdsweb/standards/imdb/imdb-menu.htm>.

3 Coverage and frames

3.1 Scope and purpose

The target population is the set of units about which information is wanted and estimates are required. Practical considerations can dictate that a survey population be defined which excludes some units in the target population or which is comprised of differently defined units through which the target population can be accessed.

A frame is any list, material or device that delimits, identifies, and allows access to the elements of the survey population. Frames are generally of two types: area frames and list frames. A list frame is a list of units in the survey population. Area frames are usually made up of a hierarchy of geographical units which in turn contain units in the survey population; that is, the frame units at one level can be subdivided to form the units at the next level. All of the elements included in the frame constitute the frame population. Frames are often much more than a simple list of units or a map with geographic units delineated. A frame usually includes other information (e.g., identification, contact, classification, address, size, maps in case of geographical units) to be used in carrying out the survey.

Coverage is the completeness of the information for the target population that would be derived if all of the frame units were to be surveyed. Coverage errors are discrepancies in statistics for the target population versus those for the frame population. These errors are a function of both the frame undercoverage (overcoverage) of the target population and of coverage errors occurring during survey operations resulting in differences in the survey estimate for those actually covered from those for which an estimate was required. Coverage errors can have both spatial and time dimensions.

3.2 Principles

The survey population should be reasonably consistent with the target population in order for the survey results to be relevant.

In turn, the survey frame should conform to the survey population. Frame coverage errors (such as missing in scope units, included out-of-scope units, misclassified units and duplicates) can complicate the survey process resulting in cost increases, loss of timeliness and also will diminish the accuracy (from bias and variance points of view) of the estimates.

Frame data should be up-to-date and accurate because of their use in stratification, sample selection, collection, follow-up, data processing, imputation, estimation, record linkage, quality assessment and analysis. Erroneous frame data are likely to bias or diminish the reliability of the survey estimates and to increase data collection costs.

Survey designs and operations should implement procedures to minimize coverage error and its impact.

3.3 Guidelines

3.3.1 Design

- Test possible frames at the planning stage of a survey for their suitability and quality. Assess the coverage of the frame and of the target collection units.
- When no single frame can provide the required coverage of the target population, use a multiple frame methodology. A multiple frame is a combination of two or more frames such as a list and area frame or two or more list frames. Generally avoid using multiple frames unless no single existing frame is adequate. When several frames exist, some of which are incomplete but less expensive to use and others more complete but prohibitively expensive, consider the use of multiple frames.

- Consider also the use of Random Digit Dialling (RDD) for some telephone surveys, by itself or in combination with other area or list frames.
- Sometimes no cost effective frame exists for the population of units of interest. In such situations consider using multi stage or indirect sampling methods.
- At Statistics Canada, several lists are maintained for use as frames by its surveys. The Business Register is available for surveys of businesses and institutions. For agricultural surveys, the Farm Register is the usual frame. For household surveys, the Address Register, the Labour Force Survey frame (which is an area frame), and the Census of Population geographic units are options to consider. In situations where one of these recognized frames is not the best choice for addressing a survey's target population, other frames e.g. lists of immigrants or databases of importers or exporters; should be considered.
- Ensure that the frame is as up to date as possible relative to the reference period for the survey.
- Retain and store information about sampling, rotation and data collection so that coordination between surveys can be achieved and respondent relations and response burden can be better managed. For example, record how often each unit is selected by each survey that is using the frame.
- For statistical activities from administrative sources or for derived statistical activities, where coverage changes may be outside the control of the immediate manager, determine and monitor coverage, and negotiate required changes with the source manager.
- Make adjustments to the data or use supplementary data from other sources to offset coverage error of the frame.
- Where possible, use the same frame for surveys with the same target population to improve coherence, avoid inconsistencies, facilitate combining estimates from the surveys and to reduce costs of frame maintenance and evaluation.
- To create geographic frame units at Statistics Canada, use the Generalized Area Delineation System (GARDS), a partially generic auto-spatial delineation and verification system for creating the non-overlapping contiguous geographic frame units.
- When multiple frames exist, they can be used to assess the completeness of one of the frames.
- Implement survey procedures to detect and correct coverage errors from the frame. Provide feedback to update and maintain the frame.
- Implement training and procedures for data collection and data processing staff aimed at minimizing coverage error (e.g. procedures to ensure accurate confirmation of lists of dwellings for sampled area frame units).
- Design survey questionnaires and related materials so as to minimize coverage errors committed by respondents (e.g erroneous listing on a questionnaire of persons in a dwelling, omission of in scope locations in an establishment survey).

3.3.2 Maintenance

- To improve and/or maintain the level of quality of the frame, incorporate procedures to eliminate duplication and to update for births, deaths, out-of-scope units and changes in characteristics.
- Incorporate frame updates in the timeliest manner possible.
- Minimize frame errors through effective training of staff, an emphasis on the importance of coverage, and the implementation of quality assurance procedures for frame-related activities.
- For area frames, implement map checks to ensure clear and non-overlapping delineation of the geographic areas used in the sampling design (e.g., through field checks or the use of other map sources). When appropriate, use the information from the address register to verify field listing of residential addresses for undercoverage or overcoverage.

- Review and improve the identification of the target units missed or wrongly coded and put in place procedures to minimize this problem.
- Put in place procedures to detect and minimize errors of omission and misclassification that can result in under coverage, and to detect and correct errors of erroneous inclusion and duplication resulting in over coverage.

3.3.3 Documentation

- Include definitions of the target and survey populations, any differences between the target population and the survey population, as well as the description of the frame and its coverage errors in the survey documentation.
- Report known gaps between key user needs and survey coverage

3.4 Quality indicators

Main quality elements: accuracy, relevance

Coverage errors arise from both undercoverage and overcoverage. In addition to target population versus survey population differences, the former occur when units are erroneously omitted from the frame file while the latter occur when units are incorrectly included on the frame file (e.g. dead units). Classification errors – in industry for example – result in coverage error on the frame; undercoverage in the “correct” classification and overcoverage in the “incorrect” classification. Classification information can be used in stratification, sample selection, data processing, imputation, estimation, record linkage, and quality assessment and analysis. Contact information can be used in data collection and follow-up. Frame imperfections such as coverage errors and out-of-date characteristics are likely to bias or diminish the reliability of the survey estimates and to increase data collection costs.

Under and over coverage can undermine both the relevance and the accuracy of survey results.

- Monitor the frame quality by periodically assessing its coverage and the quality of the information on the characteristics of the units. Many techniques exist for this purpose:
 - matching the frame or a sample of the frame with comparable alternative sources, often provided by administrative records, for the survey population or subsets of it;
 - analyzing survey returns for duplicates, deaths, out-of-scope units, and changes in characteristics;
 - using specific questions on the questionnaire to aid in monitoring coverage and classification information; verifying with local authorities (e.g., regional offices, field survey staff, the survey units themselves);
 - verifying the frame or subsets of it in the field (which could include verification of out-of-scope units);
 - comparing the frame with a sample of units from a corresponding area frame;
 - updating the frame to determine changes to it;
 - checking the consistency of counts with other sources or with data from specially designed replicates;
 - using evaluative information obtained from other surveys with the same frame (Lessler and Kalsbeek, 1992).
- Monitor the frame between the time of sample selection and the survey reference period.
- Define and compare the target population and the survey population.

- Measure coverage error in censuses via Post-enumeration (Hogan, 2003) or Reverse Record Check surveys (Statistics Canada, 2004) and associated studies of overcoverage. For censuses of population and housing, coverage error can also be assessed via comparison of census counts to demographic estimates. Provide estimates not only of net error but also of components of error.
- Provide estimates of coverage error or slippage rates both for current users of survey estimates and also for designers of future iterations of the same survey.

References

Archer, D. 1995. "Maintenance of business registers." *Business Survey Methods*.

B.G. Cox et al (eds.) New York. Wiley-Interscience, p. 85-100.

Hartley, H.O. 1962. "Multiple frame surveys." *Proceedings of the Social Statistics Section*. American Statistical Association. p. 203-206.

Hogan, H. 2003. "The Accuracy and Coverage Evaluation: Theory and Design." *Survey Methodology*, Vol. 29, no. 2. p. 129-138.

Kott, P.S. and F.A. Vogel. 1995. "Multiple-frame business surveys." *Business Survey Methods*, B.G. Cox et al (eds.) New York. Wiley-Interscience, p. 185-203.

Laniel, N. and H. Finlay. 1991. "Data quality concerns with sub-annual business survey frames." *Proceedings of the Section on Survey Research Methods*. American Statistical Association. p. 202-207.

Lessler, J.T. and W.D. Kalsbeek. 1992. *Nonsampling Errors in Surveys*. New York. Wiley-Interscience. 432 p.

Massey, J.T. 1988. "An overview of telephone coverage." *Telephone Survey Methodology*. R.M. Groves et al (eds.) New York. Wiley-Interscience. p. 3-8.

Statistics Canada. 2003. *Survey Methods and Practices*. Statistics Canada Catalogue no. 12-587E-XPE. Ottawa, Ontario. 396 p.

Statistics Canada. 2004. 2001 *Census Coverage Technical Report*. Statistics Canada Catalogue no. 92-394-XIE. Ottawa, Ontario. 85 p.

Statistics Canada. 2008. *Methodology of the Canadian Labour Force Survey*. Statistics Canada Catalogue no. 71-526-X. 116 p.

Swain, L., J.D. Drew, B. Lafrance and K. Lance. 1992. "The creation of a residential address register for coverage improvement in the 1991 Canadian Census." *Survey Methodology*. Vol. 18, no. 1. p. 127-141.

4 Sample design

4.1 Scope and purpose

Sampling is a means of selecting a subset of units from a target population for the purpose of collecting information. This information is used to draw inferences about the population as a whole. The subset of units that are selected is called a sample. The sample design encompasses all aspects of how to group units on the frame, determine the sample size, allocate the sample to the various classifications of frame units, and finally, select the sample. Choices in sample design are influenced by many factors, including the desired level of precision and detail of the information to be produced, the availability of appropriate sampling frames, the availability of suitable auxiliary variables for stratification and sample selection, the estimation methods that will be used and the available budget in terms of time and resources.

4.2 Principles

There are two types of sampling: non-probability and probability sampling. Non-probability sampling uses a subjective method of selecting units from a population, and is generally fast, easy and inexpensive. Therefore, it's sometimes useful to perform things like preliminary studies, focus groups or follow up studies. However, in order to make inferences about the population, one must make the often false assumption that the sample is representative. Probability sampling is based on three basic principles that make up the statistical framework. First, it is based on randomization, i.e. the units in the sample are selected at random. Second, all survey population units have a known positive probability of being selected in the sample, and third, we can calculate those probabilities, which are then used to calculate estimates along with estimates of the sampling error. The ability to make reliable inferences about the entire population and to quantify the error in the estimates makes probability sampling the best choice for most statistical programs.

The sample design should be as simple as possible. The aim is to produce estimates that are both precise and accurate enough to meet survey requirements. Precision is measured by the variance of an estimator. Lack of accuracy manifests itself through biases which are often introduced via nonsampling factors such as inaccurate reporting, processing, and measurement, as well as errors from non-response and incomplete reporting.

4.3 Guidelines

4.3.1 Design

- When determining sample size, take into account the required levels of precision needed for the survey estimates, the type of design (e.g., clustering, stratification) and estimator to be used, the availability of auxiliary and contact information, budgetary constraints, as well as other factors, such as non-response, presence of out-of-scope units, attrition in longitudinal surveys, etc. For periodic surveys, take into account expected births and deaths of units within the changing survey population. It's worth noting that the precision of survey estimates is usually influenced more by the total sample size than by the sampling fraction (ratio of the sample size to the population size).
- It is important to remember that most surveys produce estimates for many different variables, and optimizing the sample for one particular variable may have detrimental effects on other important variables. Handle this problem by first identifying the most important variables and then using this subset of variables to determine the sampling strategy to be adopted, which often requires a compromise between optimal strategies for the variables in the subset. See Bethel (1989).

- Stratification consists of dividing the population into subsets (called strata). Within each stratum, an independent sample is selected. The choice of strata is determined by the objectives of the survey, the availability of variables on the frame, the distribution of the variable of interest, and the desired precision of the estimates. Most surveys produce estimates for various domains of interest (e.g., provinces). If feasible, take this into account in the design by stratifying appropriately (e.g., by province). Otherwise, it will be necessary to consider special methods at the estimation stage to produce estimates for these domains (see Section 2.10). To achieve statistical efficiency, create strata in such a way that each stratum contains units that are as homogeneous as possible with respect to the information collected in the survey. For longitudinal surveys, choose stratification variables that correspond to characteristics that are stable over time.
- Conduct studies to evaluate alternative sampling methods, stratification options and allocation schemes. The usefulness of these studies depends on the availability and vintage of data used to conduct the studies, whether from previous censuses, surveys or administrative data and their relation to the variables of importance to the survey. See Kish (1988).
- Establish an expected response rate using a pre-test or data from previous occasions of the same or similar surveys. This rate can in turn be used in sample size determination. A sample can be divided into waves and additional waves of sample can be released as needed based on the achieved sample by stratum. For longitudinal surveys, expected cumulated attrition for the given number of cycles must be used.

4.3.2 Methods

- For highly skewed populations, create a stratum of large units to be included in the survey with certainty (the so-called take-all stratum). These large units would normally account for a significant portion of the population totals. In order to reduce respondent burden, the creation of a stratum of very small units to be excluded from the surveyed population is sometimes appropriate. See Baillargeon et al. (2007). It is important to distinguish between the non-surveyed portion of the survey population (the take-none stratum) which is part of the survey population but not sampled and the out-of-scope units which are not part of the survey population. The contribution for the take-none stratum may be estimated using models.
- Sometimes the information needed to stratify the population is not available on the frame. In such cases, a two-phase (or double) sampling scheme may be used, whereby a large sample is selected in the first phase to obtain the required stratification information. This first sample is then stratified and in the second phase, a subsample is selected from each stratum within the first-phase sample. Consider the cost of sampling at each phase, the availability of the information required at each phase, and the gain in precision obtained by stratifying the first-phase sample.
- In practice, it is sometimes not feasible to directly select and contact the units that will report the requested information, due to either cost or lack of information. In such cases, a two-stage sampling scheme may be used by first selecting clusters (called primary sampling units) of reporting units, and then further subsampling within each of the selected primary sampling units to obtain a sample of reporting units. Budgetary or other constraints may necessitate more than two stages (a multi-stage design). Determine how many stages of sampling are needed and which sampling units are appropriate at each stage. For each possible type of units, consider the availability of a suitable frame of such units at each stage or the possibility of creating such a frame for the survey, ease of contact and of data collection/measurement, the quality of the data provided by the units, and the cost of collection. Multi-stage designs are by definition clustered sample designs. Clustering reduces data collection cost but may result in increased variances due to intra-cluster correlation.
- If samples are to be selected from two or more frames, care must be taken in dealing with units that belong to more than one frame. It is important that the frame membership of each unit can be determined. The principle that the sample design should be simple is especially important when multiple frames are used. Choose sample designs that will simplify estimation procedures.

- In determining sample allocation and size for stratified samples, account for expected rates of misclassification of units and other deficiencies on the frame. If not properly considered at the sampling stage, survey estimates will not be as precise as planned. Address this problem at the estimation stage (see Section 2.10).
- In some complex sample designs, one must establish a design effect (DEFF) in order to determine the sample size. Results from previous or similar surveys should be used in calculating DEFF. See Gambino (2001), Kish (1965) and Gabler et. al. (2006).
- For more complex situations, for example, when surveying rare or mobile populations, or when the sampling frame is a list of units that are linked, but don't correspond directly, to the units of the target population, special designs may be necessary. Such techniques as indirect sampling, network sampling or adaptive cluster sampling, to name a few, may need to be considered. See Lavallée (2007) and Thompson and Seber (1996).
- Random Digit Dialling (RDD) is a technique that has been popular for some types of household survey. RDD schemes have the potential for bias because there are households that do not have a landline telephone. With the increasing prevalence of cell-phone-only households, this problem will become more acute unless RDD starts being used for cell phone numbers as well. The potential biases need to be studied carefully before choosing the RDD option for a survey.

4.3.3 Periodic surveys

- For periodic surveys that use designs in which the sample size grows as the population increases, it is often appropriate to develop a method to keep the sample size and therefore collection costs, stable. One approach is to use a random drop method to maintain a stable sample size over time.
- For periodic surveys, make the design as flexible as possible to deal with future changes, such as increases or decreases in sample size, restratification, resampling and updating of selection probabilities. If estimates are required for specified domains of interest (e.g., subprovincial estimates), form the strata by combining small stable units related to the identified domains (e.g., small geographical areas), if possible. Future changes in definitions of the strata will then be easier to accommodate.
- For periodic surveys, if efficient estimates of change are required or if response burden is a concern, use a rotation sampling scheme that replaces part of the sample in each period. The choice of the rotation rate will be a compromise between the precision required for the estimates of change, and the response burden on the reporting units. Lowering the rotation rate will increase the precision of the estimates of change, but may lower the response rate over time because of increased response burden. A low rotation rate has the additional benefit of reducing costs if the first contact is substantially more expensive than subsequent contacts.
- For periodic surveys, develop procedures to monitor the quality of the sample design over time. Set up an update strategy for the selective redesign of strata that have suffered serious deterioration because of uneven growth.

4.3.4 Longitudinal surveys

- For longitudinal panel surveys, determine the length of the panel (its duration of time in the sample) by balancing the need for duration data with sample attrition and conditioning effects. Use a design with overlapping panels (i.e., with overlapping time span) when there is a need to produce cross-sectional estimates along with the longitudinal ones.
- It is particularly important to choose simple design features, i.e., one single frame, and as few stages or phases as possible as the estimation procedures becomes extremely complex with the number of waves.
- It is recommended that longitudinal surveys should be designed primarily to produce longitudinal estimates. Trying to satisfy both cross-sectional and longitudinal requirements can result in very complex design and estimation procedures. If production of cross-sectional estimates is required, the use of a "top up" sample to cover births and new immigrants is recommended.

4.3.5 Implementation

- At the implementation stage, compare the size and characteristics of the actual sample to what was expected. Compare the precision of the estimates to the planned objectives. Reassess the assumptions used at the design stage. For example, evaluate non-response (non-contacts, refusals, etc.) and calculate design effects.
- Where possible, use generalized sample selection software instead of tailor-made systems. One such system is the Generalized Sampling System (GSAM) developed by Statistics Canada. GSAM is especially useful for managing sample selection and rotation for periodic surveys. By using generalized systems, one can expect fewer programming errors, as well as some reduction in development costs and time.

4.3.6 Documentation

- Write detailed and thorough documentation of all aspects of the sample design: which frames were chosen and why, how units were formed, how they were stratified, how sample size was determined and how sample was allocated, choice of stages and/or phases, which sampling schemes were used and why, and so on.

4.4 Quality indicators

Main quality element: accuracy

In addition to the points below, the reader should consult Statistics Canada Policy on informing users of data quality and methodology (http://icn-rci.statcan.ca/10/10c/10c_010_e.htm), which includes information that is pertinent here, particularly section E.1, subsection 2.3.

- Provide measures of the representativity of the sample: overcoverage and undercoverage, exclusions, comparisons to external sources (e.g., compare external demographic totals to those obtained from the survey).
- Compare the sample size observed to the expected sample size. This is especially important for multistage surveys where the sample size below the first stage may be hard to predict accurately.
- Compare response, attrition and out-of-scope rates to those that were assumed at the planning stage.
- Provide measures of sampling error: produce variances and/or CVs and compare them to the values expected at the planning stage; if design effects were used at the planning stage, compare them to realized design effects.
- For variables that were used to stratify the frame and/or allocate the sample, compare their observed CVs to their target values from the design stage.
- If possible, compare the observed homogeneity of strata to their homogeneity when they were formed; for a repeated survey, track the deterioration of the strata over time. Measure the frequency of stratum jumpers and of classification errors.

References

- Bethel, J. 1989. "Sample allocation in multivariate surveys." *Survey Methodology*. Vol. 15, no. 1. p. 47-57.
- Cochran, W.G. 1977. *Sampling Techniques*. New York. Wiley. 428 p.
- Gambino, J. 2001. *Design Effect Caveats*. Internal document. Statistics Canada.
- Gabler, S., S. Hader and P. Lynn. 2006. "Design Effects for Multiple Design Samples." *Survey Methodology*. Vol. 2, no. 1. p. 115-120.
- Hidiroglou, M.A., 1994. "Sampling and estimation for establishment surveys: stumbling blocks and progress." *Proceedings of the Section on Survey Research Methods*. American Statistical Association. p. 153- 162.
- Hidiroglou, M.A. and K.P. Srinath. 1993. "Problems associated with designing sub annual business surveys." *Journal of Business and Economic Statistics*. Vol. 11. p. 397-405.
- Kalton, G. and C.F. Citro. 1993. "Panel surveys: adding the fourth dimension." *Survey Methodology*. Vol. 19, no. 2. p. 205-215.
- Kish, L. 1965. *Survey Sampling*. New York. Wiley-Interscience. 664 p.
- Kish, L. 1988. "Multi-purpose Sample Designs." *Survey Methodology*. Vol. 14, no. 1. p. 19-32.
- Lavallée, P. 2007. *Indirect Sampling*. New York. Springer. 256 p.
- Lohr, S. 1999. *Sampling: Design and Analysis*. California. Duxbury Press, 512 p.
- Särndal, C.-E., B. Swensson and J. Wretman. 1992. *Model Assisted Survey Sampling*. New York. Springer-Verlag. 694 p..
- Statistics Canada. 2008. *Methodology of the Canadian Labour Force Survey*. Statistics Canada Catalogue no. 71-526-X. Ottawa, Ontario. 116 p.
- Statistics Canada. 2003. *Survey Methods and Practices*. Statistics Canada, Catalogue no. 12-587-XPE. Ottawa, Ontario. 396 p.
- Thompson, S.K. and G.A. Seber. 1996. *Adaptive Sampling*. New York. John Wiley and Sons. 288 p.
- Tillé, Y. 2001. *Théorie des sondages – Échantillonnage et estimation en populations finies*. Paris, Dunod.

5 Questionnaire design

5.1 Scope and purpose

A questionnaire is a set of questions designed to gather information from a respondent. It is the interface between the respondent and the researcher, and therefore plays a central role in the data collection process. A questionnaire may be interviewer-administered or respondent-completed, using different methods of data collection.

5.2 Principles

Questionnaires play a central role in the data collection process and influence the image of a statistical agency. They have a major impact on respondent behaviour, interviewer performance, collection cost and respondent relations and therefore on data quality.

A well-designed questionnaire should collect data that correspond to the survey's Statement of Objectives while taking into account the statistical requirements of data users, administrative and data processing requirements as well as the nature and characteristics of the respondent population. Good questionnaires impose low response burden and remain both respondent and interviewer-friendly. The question design and wording must encourage respondents to complete the questionnaire as accurately as possible.

To this end, the questionnaire must focus on the topic of the survey, be as brief as possible, flow smoothly from one question to the next and facilitate respondents' recall. Moreover, well-designed questionnaires should facilitate the coding and capture of data. They should minimize the amount of edit and imputation that is required, and lead to an overall reduction in the cost and time associated with data collection and processing. For more information, refer to the "Policy on the Review and Testing of Questionnaires" (Statistics Canada, 2002a).

5.3 Guidelines

5.3.1 Informing respondents

- It is the policy of Statistics Canada to provide all respondents with information about: the purpose of the survey (including the expected uses and users of the statistics to be produced from the survey), the authority under which the survey is taken, the collection registration details, the mandatory or voluntary nature of the survey, confidentiality protection, the record linkage plans and the identity of the parties to any agreements to share the information provided by those respondents. For more information, refer to the "Policy on Informing Survey Respondents" (Statistics Canada, 1998).

5.3.2 Relevance

- Consulting with data users during the questionnaire design process allows for clear understanding of how the data are to be used. It is important to undertake a review of existing subject matter literature and surveys, nationally and internationally, before designing a new questionnaire. This should allow for a well-designed questionnaire that meets the users' needs.

5.3.3 Content and wording

- The opening questions should be applicable to all respondents, be easy and interesting to complete, and establish that the respondent is a member of the survey population.
- Use words and concepts in questionnaires that have the same meanings for both respondents and questionnaire designers, and, in the case of businesses, choose questions, time reference periods, and response categories that are compatible with the establishment's record-keeping practices.
- Choose question design and wording that encourage respondents to complete the questionnaire as accurately as possible. The questionnaire must focus on the topic of the survey, be as brief as possible, flow smoothly from one question to the next, facilitate respondents' recall and direct them to the appropriate information source (Converse and Presser, 1986 and Fowler, 1995).

5.3.4 Coherence

- To the extent possible, harmonize concepts and wording with those already in use. When appropriate, reuse questions from other surveys.
- Verify French and English versions of the questionnaire for consistency.
- All members of the project team should be involved in the review of the questionnaire since each can comment from a different perspective. Team members can provide insight into whether the proposed questionnaire is conducive to good quality survey data, straight-forward programming (in computer assisted environments), and efficient post-collection data processing. More specifically, team members can evaluate the complexity of the questions and flow of the questionnaire, the impact of question structures on respondent behaviour and the detail of the questions relative to the sample size and analytical plan.

5.3.5 Layout

- Design self-completed questionnaires to be attractive and easy to complete. To this end, give a positive first impression in the cover letter and front cover, and make the questionnaire appear professional and businesslike. If it is to be interviewer-administered, make the questionnaire interviewer-friendly.
- To minimize the possibility of reporting errors, ensure that the instructions to respondents and/or interviewers are short, clear, and easy to find. Provide definitions at the beginning of the questionnaire or in specific questions, as required. Ensure that time reference periods and units of response are clear to the respondent, use boldface print to emphasize important items, specify “include” or “exclude” in the questions themselves (not in separate instructions), and ensure that response categories are mutually exclusive and exhaustive.
- With respect to the questionnaire layout, provide titles or headings for each section of the questionnaire, and include instructions and answer spaces that facilitate accurate answering of the questions. Use colour, shading, illustrations and symbols to attract attention and guide respondents or interviewers to the parts of the questionnaire that are to be read and to indicate where answers are to be placed. At the end of the questionnaire, provide space for additional comments by respondents and include an expression of appreciation to the respondent (Converse and Presser, 1986 and Fowler, 1995).

5.3.6 Data collection

- Carefully consider and evaluate different modes of data collection when designing a questionnaire. Be aware of the pros and cons of newer methods such as electronic or internet data reporting. The collection mode has implications on the amount of detail, complexity, and number of questions that can be asked of a respondent, as well as the sensitivity of the subject matter being requested.
- Build awareness among survey designers and data analysts that the mode of collection has an influence on the quality and measurement of the information collected.
- Use the optimal rules for each data collection mode when designing the questionnaire. The use of open and closed questions, mark one or mark all that apply responses, the use of rankings and ratings, as well as the question and response order can all significantly impact on respondent behaviour (De Leeuw, 2005 and Dillman and Christian, 2003).

5.3.7 Testing and evaluation

- Choose among a wide range of methods to test and evaluate the questionnaire. This could include qualitative tests such as focus groups or cognitive tests, pre-tests or pilot tests. The suitability and intensity of their use depend on various factors and circumstances. These include the type and size of the survey, the survey’s content, utilization of previous survey questions or standard questions, whether it is an ongoing collection or not, the method of data collection, the project schedule, the budget, and the availability of resources. Multiple reviews may be necessary and this will impact the cost and project schedule (Couper, Lessler, Martin, Martin, Presser, Rothgeb, and Singer, 2004).

5.4 Quality indicators

Main quality elements: accuracy, relevance, coherence.

- Measurement error is the difference between measured values and true values. It consists of bias (systematic error introduced where the measuring instrument is inaccurate – this error remains constant across survey replications) and variance (random fluctuations between measurements which, with repeated samples, would cancel each other out). Sources of measurement error are the survey instrument, mode of data collection, respondent's information system, respondent, and interviewer.
- A description of the processes developed to reduce measurement error associated with the survey instrument and to optimize the comparability of the data being collected should be made available. This will indicate to users the accuracy and reliability of the measures as well as the coherence of the data being collected with other statistical information. These processes might include questionnaire development, pilot studies, questionnaire testing, interviewer training, etc.
- The accuracy of the data as well as its coherence may be influenced in many ways. For example:
 - by the sensitive nature of the information sought
 - if the words and concepts used do not have the same meaning for both respondents and survey takers
 - if concepts and wording are not harmonized with terms already in use, particularly for business surveys

If so, an assessment of the direction and amount of bias of these items should be made to evaluate the impact on data quality.

- Making questionnaires available to users will help them assess the relevance and coherence of the data, considering their own needs and the other data sources available to them.
- Users should be kept informed of any modification to the questionnaire over time and an assessment of the impact of such changes on data comparability should be performed.
- Documentation on the data quality of the survey should also include any problems with question wording, response burden, refusal rates or other relevant information.

References

Converse, J.M. and S. Presser. 1986. *Survey Questions: Handcrafting the Standardized Questionnaire*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-063. Thousand Oaks, California. Sage Publications. 80 p.

Couper, M. P., Judith T. Lessler, E.A. Martin, J. Martin, J.M. Rothgeb and E. Singer. 2004. *Methods for Testing and evaluating survey questionnaires*. Hoboken, New Jersey. John Wiley and Sons, Inc.

De Leeuw, Edith D. 2005. "To mix or not to mix data collection modes in surveys." *Journal of Official Statistics*. Vol. 21, no. 2. p 233-255.

Dillman, Don A. 2000. *Mail and Internet Surveys: The Tailored Design Method*. 2nd ed. Toronto. John Wiley and Sons, Inc.

Dillman, Don A. and Leah M. Christian. 2003. "Survey Mode as a source of instability in responses across surveys." Presented at the *Workshop on stability of methods for collecting, analyzing and managing panel data*, American Academy of Arts and Science, Cambridge, Massachusetts. March 2003.

Fowler, F.J. Jr. 1995a. *Improving Survey Questions: Design and Evaluation*. Applied Social Research Methods Series, 38. Thousand Oaks, California. Sage Publications. 200 p.

Statistics Canada. 1998. "Policy on Informing Survey Respondents." *Statistics Canada Policy Manual*. Section 1.1. Last updated March 4, 2009.
http://icn-rci.statcan.ca/10/10c/10c_001_e.htm.

6 Data collection, capture and coding

6.1 Scope and purpose

Data collection is any process whose purpose is to acquire or assist in the acquisition of data. Collection is achieved by requesting and obtaining pertinent data from individuals or organizations via an appropriate vehicle. The data is either provided directly by the respondent (self-enumeration) or via an interviewer. Collection also includes the extraction of information from administrative sources which may require asking the respondent permission to link to administrative records.

Data capture refers to any process that converts the information provided by a respondent into electronic format. This conversion is either automated or involves staff keying the collected data (keyers). Data coding is any process that assigns a numerical value to a response. Coding is often automated, however, more complex decisions usually require human intervention (coders).

Often, survey operations involve a high degree of automation, which leads to the availability of paradata, information related to a survey process. Examples of paradata include an indicator of whether or not a unit is in the sample; history of calls and visits; trail of key strokes (audit trail); mode of collection; administrative information (e.g. interviewer profile); and cost information.

Data collection is not only the source of information; it is also the main contact a survey-taking agency has with the public who needs to be convinced to participate. Data capture and coding produce the formatted data used as input by all the subsequent survey processes. Data collection, data capture and coding operations often use a large portion of the survey budget, requiring considerable human and physical resources, as well as time.

6.2 Principles

Respondents are a survey-taking organization's most valuable resource. Every variable that cannot be derived from other existing sources is a burden to the respondent. The amount of time and energy a respondent spends on providing data must be minimized. Privacy and security must be respected throughout all data gathering and processing operations. Given that these operations have a high impact on data accuracy, quality and performance measurement tools should be used to manage data collection, data capture and data coding processes.

6.3 Guidelines

6.3.1 Data collection

- Careful planning of the collection process should include the establishment of roles and responsibilities regarding all aspects linked to collection, including its communication strategy, execution, assessment, monitoring, contingency planning and security.
- Design the collection process in order to reduce respondent burden and collection cost, and to maximise timeliness and data accuracy. Data could be collected through self-enumeration, telephone interviews or in-person interviews with either a paper or an electronic questionnaire (e.g. electronic data reporting, Internet, computer assisted interviewing). To achieve the design objectives stated above more easily, consider using more than one method throughout the collection cycle. For instance, collection may start with self-enumeration using a paper or Internet questionnaire and may finish with an in-person interview. For self-enumeration surveys, use multiple events (e.g. pre-collection advertisement card, introduction letter with the questionnaire, reminder card, reminder call or visit) over the collection period in order to stimulate the return of questionnaires. Examine whether some of the data elements

could be acquired via administrative records instead of the more costly and sometimes less accurate traditional collection methods. Consider conducting the collection as a supplement to a large scale survey. This would not only potentially reduce survey costs and respondent burden, but also make available a wealth of information for non-response adjustment. When feasible, conduct pilot studies or tests to help determine or fine-tune the collection operation.

- Establish appropriate sample control procedures and measures for all data collection operations (e.g. delivery and return of paper questionnaires, follow-up for gaps or inconsistencies, follow-up for non-response). Such procedures track the status of sampled units from the beginning through to the completion of data collection so that data collection managers and interviewers can assess progress at any point in time. This is particularly important for surveys that use many data collection modes and that move cases from one mode to another (or from one collection centre to another). Sample control procedures are also used to ensure that every sampled unit is processed through all the steps subsequent to data collection (i.e. capture and coding steps), with a final status being recorded. Sample control measures can be used to evaluate the efficiency of those procedures.
- Establish and maintain good respondent relationships in order to obtain a good response rate. Such measures can include advertising the upcoming survey; an introductory letter to inform the respondents that they will be part of a survey; an informative brochure with key statistics to maintain their interest in participating in the survey (in particular for longitudinal surveys); or procedures facilitating access to publicly available information, for example on a website, a guide to complete the questionnaire, or helpline (in particular for self-enumeration surveys); or a letter thanking them for their participation. These measures will help to sensitize the units selected in the sample to participate in the survey.
- When collecting data, ensure that the respondent or the appropriate person within the responding household or organization is contacted at the appropriate time. Allow the respondent to provide the data in a method and format that is convenient to the individual and his or her organization. This will help increase response rates and improve the quality of the information obtained from the respondents. Special reporting arrangements should be considered in specific cases in order to reduce respondent burden and to facilitate the collection of information. For example, consider creating a special collection arrangement for enterprises that are involved in many surveys. For households, when the targeted respondent is not available, establish rules to determine who could act as an appropriate proxy, should this be an option.
- For collection by interview, determine the best time to call or to visit survey units based on paradata acquired during previous iterations of the survey or from a similar survey. Manage calls or visits in such a way that respondents are contacted at the best time and that the number of call or visit attempts does not exceed a useful maximum. In addition, respondents should each be assigned a priority level so that they may be contacted or visited for interviews based on order of importance. Assignment priority should be based on the target effective sample size by domain of interest that would lead to estimates accurate enough (having low bias and variance) to be released. For business surveys, this would mean giving higher priority to large or influential units first, possibly at the risk of missing smaller units. For household surveys, priority should be given to units less likely to respond. A score function is a useful tool for prioritization. For telephone interviews, use an automated system to manage case call scheduling. Such a system should also prioritize cases.
- Interviewers are vital to the success of data collection operations. Interviewer manuals and training must be carefully prepared and planned, since they provide the best way to guarantee data quality (e.g. high response rate and accurate responses), the comprehension of survey concepts and subject matter, as well as to ensure proper answers to questions from respondents. Training can use different approaches such as home study, classroom training, mock interviews or live interviews. Interviewing skills of interviewers should be monitored to ensure that they conform to a pre-established list of standards (e.g. reading questions as written in the questionnaire). This monitoring should also be used to identify

strengths and weaknesses in the interviewer's skill set, to provide feedback to the interviewers and to focus training on weaker areas. Depending on the interviewing mode and resources, the monitoring may either be done using recordings of the interviews or live. Consultation with interviewers and staff directly responsible for collection operations will help in the development of better training tools. Follow-up interviews with respondents may also be used to get the respondent's point of view on how the interview was carried out.

- Tracing should be conducted to locate and contact respondents when the available contact information on the survey unit is likely to be outdated. Tracing increases response rate and also helps in determining if the sampled unit is still in scope. Consider using administrative sources (e.g. telephone files, other survey frames) prior to survey collection and during collection in order to update contact information. During collection, facilitate high-quality tracing by obtaining extra information related to the sample unit, for example, the names of other family members, relationship, age, etc. Local knowledge might also be useful. Consider forming a team of tracing experts when the survey is repeated or its collection period is over several months. In between cycles, facilitate feedback from the respondent to update contact information. For example, provide the respondents with a "change of address" card and ask them to notify the Agency if a move occurs. Collect tracing information (e.g. internet address, cell phone number) that can be used in the subsequent survey cycles.
- For self-enumeration surveys, once the data is received, verify gaps or inconsistencies related to accuracy of the coverage information and the quality of the data provided. Follow-up interviews may be needed in some cases (e.g. when the questionnaire is missing a large number of items). Assign a follow-up priority based on the statistical importance of these units and of the missing items.
- Given that self-enumeration surveys tend to result in lower unit response rates, consider following up with non-respondents by telephone or in person to obtain their participation or conduct an interview. Ensure that collection staff is informed in a timely fashion of the registration of returned questionnaires in order to avoid unnecessary follow-ups. This type of follow-up is particularly important in the case of longitudinal surveys where the investment is clearly more long-term and the sample is subject to accumulating attrition (and possibly bias) due to non-response at each survey occasion. Unit non-response follow-ups should also be prioritized with the approach described above for managing interview surveys. Paradata (e.g. number of call or visit attempts) can also be useful to prioritize follow-ups.
- As a last step of the collection operation, consider contacting a sub-sample or all of the non-responding units (including unresolved cases) to determine whether they are in scope or not (e.g. active business or not, occupied dwelling or not); and, if so, a critical data item such as size (e.g. business total income, household size) should be obtained. This information will be useful for the non-response adjustment. In some instances, the information can be obtained from or approximated by current administrative data for all of the non-responding units.
- Provide plans and tools to actively manage survey data collection while it is in progress. Productivity measures (e.g. daily and cumulative number of units resolved) and cost indicators (e.g. daily and cumulative interviewer hours and travel expenditures) can be used to assess the relationship between the collection effort and the results (e.g. unit response rate). Compared with planned values, these indicators also help survey managers in their decision-making throughout the collection period. Used in conjunction with the daily response rate, the daily productivity rate and daily average unit cost provide the marginal cost of response rate increase during the course of collection. Activity and cost indicators (according to selected unit or completed questionnaire) also make it possible to evaluate the additional costs and effort required to increase response rates, particularly towards the end of the collection period.
- Every effort should be made to ensure the confidentiality of the data. Staff handling confidential data must be familiar with best practices regarding the printing, handling and filing of paper documentation, the handling of electronic files, and the rules regarding the dissemination of information.

- Consider implementing a re-interview program to assess the overall accuracy of interviewing operations.
- Use paradata to identify operational efficiency and cost-efficiency opportunities (e.g. sequence of calls, best time to call, optimal limit for calls or visits, etc.) in order to improve current and future collection processes and practices. For example, use average and distribution of interview duration to plan the next survey cycle. Interview duration can also be used to evaluate part of respondent burden. If interview duration is analysed by interviewer, it can be used as well to identify those potentially requiring additional training (e.g. those with outlying average duration).

6.3.2 Data capture

- Design the capture process in order to reduce capture cost and to maximise timeliness and data accuracy. Data items could be captured during survey collection by the respondents (e.g. Internet, EDR) or the interviewers (e.g. CATI, CAPI). This obviously reduces the cost of capture, increases the timeliness and has the potential of improving accuracy through edit rules being integrated into the computer application. When it is not feasible to integrate capture with collection, the capture is performed either by operators (manual key entry) or in an automated fashion (scanning followed by Intelligent Character Recognition). The latter is preferred as it reduces cost and often enhances accuracy of the data.
- For CATI and CAPI interviewers who often perform data capture and coding during collection, use standard collection tools and process (e.g. standard screens and standardized questions) to ease interviewer work and limit the risk of introducing a capture error. Integrate edit rules in the collection system to validate the entry of data items and allow for potential corrections of errors (i.e. keying error, response error, missing item) at the time of collection.
- Data capture operators are critical to the success of the capture operations. Ensure that they have appropriate training and tools. Prepare training material and procedures for the keyers and deliver training sessions. This will enhance the skills of the staff and thus ensure accurate capture of data collected. Use quality control methods to verify whether the accuracy of capture performed by operators meets the pre-established levels and provide them with feedback for improvement.
- Manual data capture from paper questionnaires or scanned images is subject to keying errors. Incorporate online edits for error conditions that the data capture operator can correct (i.e. edits that will identify keying errors). Record these cases for later review and analysis. When feasible, the manual operation should be tested prior to conducting the survey.
- For automated data capture, ensure that the questionnaire is designed to ease the scanning and the intelligent character recognition.
- When automated capture is used, some questionnaires cannot be scanned and others can be scanned but characters cannot be recognised. For damaged or badly scanned questionnaires, use a team of keyers to perform the capture.
- Systems for automated data capture by intelligent character recognition from scanned images should be tested prior to implementation. Such systems may cause relatively high rates of systematic errors in specific data items. It might be possible to improve the algorithms and their parameters to reduce the error rates. For the data items at high risk of systematic error, consider using keyers.
- Keyers should also be used to conduct a sample study assessment of the accuracy of automated capture. The results of such a study can be used to improve the process.
- Institute effective control of systems to ensure the security of data capture, transmission and handling, especially with new technologies such as cell phone and Internet data collection. Prevent loss of information and the resulting decline in quality, and potentially in credibility, due to system failures or human errors. Develop procedures for destroying the data when no longer needed.

6.3.3 Data coding

- Design the coding process in order to reduce coding cost, to maximize timeliness and data accuracy. Often data items are pre-coded during collection with the use of closed questions. This obviously reduces the cost of coding and could also improve accuracy. When this is not feasible and open questions are asked, the coding is performed after collection either by operators or in an automated fashion (e.g. using the Automated Coding with Text Recognition system). The latter is preferred as it often reduces cost and enhances accuracy of the results.
- For manual coding operations, make sure that the procedures are applied to all units of study as consistently and in as error-free a manner as possible. A computer-assisted operation is desirable. Enable the staff or systems to refer difficult cases to a small number of knowledgeable experts. Centralize the processing in order to reduce costs and make it simpler to take advantage of available expert knowledge. Given that there can be unexpected results in the collected information, use processes that can be adapted to make appropriate changes if necessary from the point of view of efficiency. When feasible, the manual operation should be tested prior to conducting the survey.
- Data coding operators are critical to the success of the coding operations. Ensure that they have appropriate training and tools. Prepare training material and procedures for the coders and deliver training sessions. This will enhance the skills of the staff and thus ensure accurate coding of data. Use quality control methods to verify whether the accuracy of coding performed by operators meets the pre-established levels and provide them with feedback for improvement.
- For automated coding, build and maintain reference files to maximize phrases recognized while minimizing errors. When automated coding is used, often a number of cases remain uncoded. The use of a team of coders is an appropriate approach to complete these cases.
- Expert coders should be used to conduct a sample study assessment of the accuracy of automated coding. The results of such a study can be used to augment and to improve the content of reference files used.

6.3.4 Quality control

- Use statistical quality control methods to assess and improve the quality of collection, capture and coding operations. Collect and analyze quality control measures and results in a manner that would help identify the major root causes of error. Provide feedback reports to managers, staff, subject matter specialists and methodologists. Use measures of quality and productivity to provide feedback at the interviewer or operator level, as well as to identify error-causing elements in the design of the operation or its processing procedures. These reports should contain information on frequencies and sources of error (see Mudryk et al, 1994, 1996 and 2002; Mudryk and Xiao, 1996). Various software tools are available to help in this regard. These include the Quality Control Data Analysis System (QCDAS) and NWA Quality Analyst (see Mudryk, Bougie and Xie, 2002).

6.3.5 Post-mortem analysis

- Conduct a post-mortem evaluation of data collection, capture and coding operations, and document the results for future use. Evaluate the processes to identify the lessons learned with the goal of improving each of its components. For that purpose, post-survey studies are often useful.
- Use subsequent survey processes to gather useful information regarding quality that can serve as signals indicating that collection, capture and coding procedures and tools may require changes for future survey cycles. For example, the editing or data analysis stages may suggest the possibility of response bias or other collection-related problems.

6.4 Quality indicators

Main Quality Element: Accuracy

The impact of data collection and capture operations (including coding) on data quality and cost is both direct and critical, as these data are the primary inputs of a survey-taking agency, and often the most important survey expenditure components. The quality of these operations thus has a very high impact on the quality of the final product, in particular, on its accuracy.

Quality measures gathered during the data collection operation enable the survey manager to make decisions regarding the need for process modification or redesign. Important quality measures include response rates, processing error rates, follow-up rates and rates of non-response by reason. When these measures are available at all levels at which estimates are produced and at various stages of the process, they can serve both as performance measures and measures of data quality.

6.4.1 Proxy rates

Report proxy rates (i.e. percentage of cases where responses are obtained from a respondent who is not the selected survey unit) as an indicator of potential response error.

6.4.2 Non-response rates

Report non-response rates as an indicator of non-response bias. Unit non-response can be decomposed in many components, for example, the interview was prevented due to non-contact, refusal, temporary absence, technical problem, language problems or the respondent's mental or physical condition. To reflect the uncertainty related to coverage, unit non-response can also be decomposed as cases resolved (i.e. in-scope status is determined) versus unresolved (i.e. in-scope status is undetermined). Report item non-response (e.g. refusal and don't know) to key questions. Item non-response rates may vary for early and late respondents (i.e. those who require more calls or visits). Both unit and item non-response rates can be reported by domain of interest to be released (in this instance it can also serve as a release criteria) and also by sub-population (e.g. large and small businesses, young and older adults) to indicate how well the effective sample represents the population. Unit non-response rate and item non-response rate can also be combined to provide global non-response rate by item. Other useful indicators are the refusal conversion rate and tracing conversion rate (in the case of erroneous or outdated initial contact information). For surveys with topics of a more sensitive nature, refusal rate at the first contact can be reported.

6.4.3 In-scope/out-of-scope error rates

When an in-depth study is conducted on how well the collection operation has classified non-responding units as in-scope/out-of-scope (e.g. business: active/inactive, dwelling: occupied/unoccupied), report the rate of being classified as in-scope when truly out-of-scope and the rate of being classified as out-of-scope when truly in-scope. These rates can be reported by domain of interest to be released.

6.4.4 Average interview length distribution

Report the average and the distribution of interview duration. In particular, report the percentage of extremely short interviews which may indicate problems with the reported data. Analysis of interview length can also be used to evaluate part of respondent burden.

6.4.5 Mode effect

A mode effect is a measurement bias that is attributable to the mode of data collection. Ideally, mode effects can be investigated using experimental designs where sample units are randomly assigned into two or more groups. Each group is surveyed using a different data collection mode. All other survey design features are controlled. Differences in the response distributions for the different groups can be compared and assessed. Other methods, such as the propensity score method or regression analysis, can be used to assess mode effects when experimental designs cannot be applied.

6.4.6 Edit reject rates

Report the rate of edit rejects, the number and type of corrections applied by domain, collection mode, processing type, data item and language of collection. This will help in evaluating the quality of the data and the efficiency of the editing function used in collection and capture operations. The edit reject rates can be decomposed by the reason of rejection, (i.e. the item is missing or the item reported is inconsistent with the normal range of values for that item or with other items reported). The latter component is an indicator of measurement error (i.e. response error + capture error).

6.4.7 Outgoing capture/coding error rates

Report outgoing capture/coding error rates in manual and automated operations calculated from results of quality verification or studies. When both manual and automated capture/coding is used, calculate composite rates. Overall rates can be calculated, as well as rates by domain, collection mode, processing type, data item and language of collection.

References

Bethlehem, J., F. Cobben, B. Schouten. 2008. "Indicators for the Representativeness of Survey Response." *Proceedings from the 2008 International Symposium on Methodological Issues, Statistics Canada*.

Couper, M.P., R.P. Baker, J. Bethlehem, C.Z.F. Clark, J. Martin, W.L. Nicholls II and J. O'Reilly (eds.) 1998. *Computer Assisted Survey Information Collection*. New York. Wiley-Interscience. 653 p.

Dielman, L. and M.P. Couper. 1995. "Data quality in a CAPI survey: keying errors." *Journal of Official Statistics*. Vol. 11, no. 2. p. 141-146.

Dillman, D. A. 2006. *Mail and Internet Surveys: The Tailored Design Method*. New York. Wiley. 554 p.

Groves, R.M. 1989. *Survey Errors and Survey Costs*. New York. John Wiley and Sons. 620 p.

Groves, R.M., P. Biemer, L. Lyberg, J. Massey, W. L. Nicholls and J. Waksberg (eds.) 1988. *Telephone Survey Methodology*. New York. Wiley-Interscience. 608 p.

Groves, R.M. and S.G. Heeringa. 2006, "Responsive design for household surveys: Tools for actively controlling survey errors and costs." *Journal of the Royal Statistical Society. Series A*. Vol. 169, no. 3. p. 439-357.

Hunter, L. and J.-F. Carbonneau. 2005. "An Active Management Approach to Survey Collection." *Proceedings from the 2005 International Symposium on Methodological Issues, Statistics Canada*.

Laflamme, F. and C. Mohl. 2007. "Research and Responsive Design Options for Survey Data Collection at Statistics Canada." *Proceedings of the Section on Survey Research Methods*. American Statistical Association.

Laflamme, F., M. Maydan and A. Miller. 2008. "Using Paradata to Actively Manage Data Collection." *Proceedings of the Section on Survey Research Methods*. American Statistical Association.

Laflamme, F. 2008. "Data Collection Research using Paradata at Statistics Canada." *Proceedings from the 2008 International Symposium on Methodological Issues, Statistics Canada*.

Laflamme, F., 2008, "Understanding Survey Data Collection Through the Analysis of Paradata at Statistics Canada." *American Association for Public Opinion Research 63rd Annual Conference, 2008. Proceedings of the Section on Survey Research Methods*. American Statistical Association.

Lepkowski, James M. et al. 2007. "Advances in Telephone Survey Methodology." *Second International Conference on Telephone Survey Methodology, Miami 2006*. Wiley series in survey methodology section. p. 363-367.

Lyberg, L., P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz and D. Trewin (eds.) 1997. *Survey Measurement and Process Quality*. New York. Wiley-Interscience. 808 p.

Mudryk, W., M.J. Burgess and P. Xiao. 1996. "Quality control of CATI operations in Statistics Canada." *Proceedings of the Section on Survey Research Methods*. American Statistical Association. P. 150-159.

Mudryk, W., B. Joyce, H. Xie. 2004. "Generalized Quality Control Approach for ICR Data Capture in Statistics Canada's Centralized Operations." *European Conference on Quality and Methodology in Official Statistics*. Federal Statistical Office, Germany.

Rosenbraum P.R. and D.B. Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika*. Vol. 70, no. 1. p. 41-45.

Statistics Canada. 1998a. "Policy on Informing Survey Respondents." *Statistics Canada Policy Manual*. Section 1.1. Last updated March 4, 2009.
http://icn-rci.statcan.ca/10/10c/10c_001_e.htm.

Statistics Canada. 2001d. *Standards and Guidelines for Reporting of Non-response Rates*. Statistics Canada Technical Report.

Statistics Canada. 2003. *Survey methods and practices*. Statistics Canada Catalogue no. 12-587-XPE. Ottawa, Ontario. 396 p.

Williams, K., C. Denyes, M. March and W. Mudryk, 1996. "Quality measurement in survey processing." *Proceedings of the 1996 International Symposium on Methodological Issues, Statistics Canada*.

7 Use of administrative data

7.1 Scope and purpose

Administrative records are data collected for the purpose of carrying out various non-statistical programs. For example, administrative records are maintained to regulate the flow of goods and people across borders, to respond to the legal requirements of registering particular events such as births and deaths, and to administer benefits such as pensions or obligations such as taxation (for individuals or for businesses). As such, the records are collected with a specific decision-taking purpose in mind, and so the identity of the unit corresponding to a given record is crucial. In contrast, in the case of statistical records, on the basis of which no action concerning an individual or a business is intended or even allowed, the identity of individuals/businesses is of no interest once the database has been finalized.

Using administrative records presents a number of advantages to a statistical agency and to analysts. Demands for statistics on all aspects of our lives, our society and our economy continue to grow. These demands often occur in a climate of tight budgetary constraints. Statistical agencies also share with many respondents a growing concern over the mounting burden of response to surveys. Respondents may also react negatively if they feel they have already provided similar information (e.g. revenue) to administrative programs and surveys. Administrative records, because they already exist, do not incur additional cost for data collection nor do they impose a further burden on respondents. Advancements in technology have permitted statistical agencies to overcome many of the limitations caused by processing large datasets. For all these reasons, administrative records are being used increasingly for statistical purposes.

Statistical uses of administrative records include (i) use for survey frames, directly as the frame or to supplement/update an existing frame, (ii) replacement of data collection (e.g. use of taxation data for small businesses in lieu of seeking survey data for them), (iii) use in editing and imputation, (iv) direct tabulation, (v) indirect use in estimation (e.g. as auxiliary information in calibration estimation, benchmarking or calendarisation), and (vi) survey evaluation, including data confrontation (e.g. comparison of survey estimates with estimates from a related administrative program).

On the other hand, one must be careful in using administrative data as there are a number of limitations to be aware of including (i) the level or the lack of quality control over the data (ii) the possibility of having missing items or missing records (an incomplete file) (iii) difference in concepts which might lead to bias problems, as well as coverage problems (iv) timeliness of the data (the collection of the data being out of the statistical agency's control, it is possible that due to external events, part or all of the data might not be received on time). Also, one must remember that there is some cost that comes with administrative data. For instance, computer systems are needed to clean and complete the data in order to make it useful. For a discussion on the advantages and disadvantages of using administrative data, see Lavallée (2000).

7.2 Principles

It is Statistics Canada's guiding principle to use administrative records whenever they present a cost-effective alternative to direct data collection. As with any data acquisition program, consideration of the use of administrative records for statistical purposes is a matter of balancing the costs and benefits. In many cases, using administrative records avoids further data collection costs and respondent burden, provided the coverage and the conceptual framework of the administrative data are compatible with the target population. In other situations, there may be costs incurred through paying for the data capture or providing some service in exchange. Depending on the use, it is often valuable to combine an administrative source with another source of information.

The use of administrative records may raise concerns about the privacy of the information in the public domain. These concerns are even more important when the administrative records are linked to other sources of data. The Policy on Informing Survey Respondents (Statistics Canada, 1998) requires that Statistics Canada provide all respondents with information such as the purpose of the survey, the confidentiality protection measures, the record linkage plans and the identity of the parties to any agreements to share the information provided by those respondents. Record linkage must be in compliance with the Agency's Policy on Record Linkage (Statistics Canada, 2008). In particular, all requests for record linkage must be submitted to the Confidentiality and Legislation Committee and approved by the Policy Committee. Requests are usually approved for specific uses only. However, in certain cases, some data requests are approved for recurring or continuous uses.

The use of administrative data may require the statistical agency to implement a number, usually only a subset, of the survey steps discussed in the other sections. This is because many of the survey steps (e.g. direct collection and data capture) are performed by the administrative organization. As a result, additional guidelines to the others presented are required to suggest ways to compensate for any differences in the quality goals of the source organization. For instance, an extensive edit and imputation program might have to be developed in order to achieve a certain level of quality required for the uses of the data.

One must keep in mind the fundamental reason for the existence of these administrative records: they are the result of an administrative program that was put in place for administrative reasons. Often, the statistical uses of these records were unknown when the program was implemented and the statistical agency often has limited impact in the development of the program. For that reason, any decisions related to the use of administrative records must be preceded by an assessment of such records in terms of their coverage, content, concepts and definitions, the quality assurance and control procedures put in place by the administrative program to ensure their quality, the frequency of the data, the timeliness in receiving the data by the statistical agency and the stability of the program over time. Obviously, the cost of obtaining the administrative records is also a key factor in the decision to use such records.

7.3 Guidelines

7.3.1 The Administrative Program

- Maintain continuing liaison with the provider of administrative records. Liaison with the provider is necessary at the beginning of the use of administrative records. However, it is even more important to keep in close contact with the supplier at all times so that the statistical agency is not surprised by any changes, and can even influence them. Feedback to the supplier of statistical information and weaknesses found in the data can be of value to the supplier, leading to a strengthening of the administrative source.
- Understand the context under which the administrative organization created the administrative program (e.g. legislation, objectives, and needs). It has a profound impact on (i) the universe covered, (ii) the contents, (iii) the concepts and definitions used, (iv) the frequency and timeliness, (v) the quality of the recorded information, and (vi) the stability over time. Pay special attention to the consistency of the concepts and data quality when there are multiple sources of administrative data, for example when each province manages its own program.
- Keep in mind that if the information provided to the administrative source can cause gains or losses to individuals or businesses, there may be biases in the information supplied which can lead to unexpected coverage problems and biases. Special studies may be needed in order to assess and understand these sources of error.

7.3.2 Assess quality

- Many of the guidelines in other sections are applicable to administrative records. Sampling and data capture guidelines will be relevant if administrative records exist only on paper and have to be coded and captured. These guidelines will also be of value for administrative data available in electronic form, including EDR (electronic data reporting). Note that these data, because they exist in electronic form, may be inherently less stable and subject to additional errors arising from data treatment and transmission processes at source. Editing and dissemination guidelines apply to all cases where a file of individual administrative records is obtained or created for subsequent processing and analysis.
- Collaborate with the designers of new or redesigned administrative systems. This can help in building statistical requirements into administrative systems from the start. Such opportunities are rare, but when they happen, the eventual statistical value of the statistical agency's participation can far exceed the time and work expended on the exercise.
- Study each data item in the administrative records that is planned to be used for statistical purposes. Investigate quality. Understand the concepts, definitions and procedures underlying collection and processing by the administrative organisation. Some of the items might be of very poor quality and thus might not be useful. For example, the quality of classification coding (e.g. occupation, industrial activity, geography) might not be sufficient for some statistical uses or might limit its use.
- Keep in mind that the longevity of the source of administrative data and its continued scope is usually entirely in the hands of the administrative organization. The administrative considerations that originally dictated the concepts, definitions, coverage, frequency, timeliness and other attributes of the administrative program may, over time, undergo changes that distort time series derived from the administrative source. Be aware of such changes, and manage their impact on the statistical program.
- Implement continuous or periodic assessment of incoming data quality. Assurance that data quality is being maintained is important because the statistical agency does not control the data collection process. This assessment may consist of implementing additional safeguards and controls (e.g. the use of statistical quality control methods and procedures, edit rules) when receiving the data, comparisons with other sources or sample follow-up studies. A good practice is to provide feedback to the administrative source to assist them in improving their data.

7.3.3 Confidentiality

- Consider privacy implications of the publication of information from administrative records. Although the Statistics Act provides Statistics Canada with the authority to access administrative records for statistical purposes, this use may not have been foreseen by the original suppliers of information (Statistics Canada, 2005). Therefore, programs should be prepared to explain and justify the public value and innocuous nature of this secondary use.
- Administrative information is sometimes used to replace a set of questions that would otherwise be asked of the respondent. In this instance, permission from the respondent may have to be obtained. Follow the Policy on Informing Survey Respondents (Statistics Canada, 1998) in this regard. When consent is not obtained, put collection procedures in place for the equivalent survey questions to be asked of the respondents.
- Administrative data often has information about specific people or businesses. Any data release from Statistics Canada is subject to the Statistics Act confidentiality provision, even when the data itself is already available in the public domain. Therefore, the guidelines for disclosure control should be considered when preparing any data analysis for release, including the release of administrative data.

7.3.4 Non-response

- Like data collected by means of a survey, administrative data are also subject to partial and total non-response. In some instances, the lack of timeliness in obtaining all administrative data introduces greater non-response. Some non-response guidelines will thus apply. Unless non-respondents can be followed up and responses obtained, develop an imputation or a weight-adjustment procedure to deal with this non-response. Administrative sources are sometimes outdated. Therefore, as part of the imputation process, give special attention to the identification of active and/or inactive units. Some imputation or transformation (e.g. calendarisation) may also be required in cases where some of the units report the data at a different frequency (e.g. weekly or quarterly) than the one desired (e.g. monthly).

7.3.5 Record linkage

- When record linkage of administrative records is necessary (e.g. for tracing respondents, for supplementing survey data, or for data analysis), conform to the Agency's Policy on Record Linkage (Statistics Canada, 2008). Privacy concerns that may arise when a single administrative record source is used are multiplied when linkage is made to other sources. In such cases, the subjects may not be aware that information supplied on two separate occasions is being combined. The Policy on Record Linkage is designed to ensure that the public value of each record linkage truly outweighs any intrusion on privacy that it represents.
- It is not always easy to combine an administrative source with another source of information. This is especially true when a common matching key for both sources is not available and record linkage techniques are used. In this case, select the type of linkage methodology (e.g. exact matching or statistical matching) in accordance with the objectives of the statistical program. When the purpose is frame creation and maintenance, or data editing, exact matching should be used. In the case of imputation or weighting, exact matching should be used, but statistical matching can be also sufficient. When the sources are linked for performing some data analyses that are impossible otherwise, consider statistical matching, e.g. matching of records with similar statistical properties (see Cox and Boruch, 1988; Kovacevic, 1999).
- When record linkage is to be performed, make appropriate use of existing software. Statistics Canada's Generalized Record Linkage Software is but one example of a number of well-documented packages.
- When data from more than one administrative source are combined, pay additional attention to reconcile potential differences in their concepts, definitions, reference dates, coverage, and the data quality standards applied at each data source. Examples are education data sources, health and crime reports, and registries of births, marriages, licenses, and registered vehicles, which are provided by various organizations and government agencies.
- Some administrative data are longitudinal in nature (e.g. income tax, goods and services tax). When records from different reference periods are linked, they are very rich data mines for researchers. Remain especially vigilant when creating such longitudinal and person-oriented databases, as their use raises very serious privacy concerns. Use the identifier with care, as a unit may change identifiers over time. Track down such changes to ensure proper temporal data analysis. In some instances, the same unit may have two or more identifiers for the same reference period, thus introducing duplication in the administrative file. If this occurs, develop an unduplication mechanism.

7.3.6 Documentation

- Document the nature and quality of the administrative data once assessed. Documentation helps statisticians decide the uses to which the administrative data are best suited. Choose appropriate methodologies for the statistical program based on administrative data and inform users of the methodology and data quality.

7.4 Quality indicators

Main quality elements: relevance, accuracy, timeliness, coherence

7.4.1 Relevance

Do the data elements that are being captured in the administrative system reflect the concepts and definitions of the data user? Although it is often less expensive to mine administrative data than to collect the information via a survey, the analytical goals must be met with the administrative data in order for it to be a useful endeavour. Indicate the source, vintage, and how well definitions and classifications match to the survey data, and to the needs of data users.

7.4.2 Accuracy

Administrative data often does not go through the same edits that survey data does. Some edits are usually performed by the administrative organization, but their nature and purposes are usually different from those of the statistical agency. As a result, data quality can be an issue when using administrative sources for statistical purposes, particularly with no or limited ability to re-contact the originator of the information. Additionally, sampled administrative data may not adhere to any standard sampling scheme, introducing possible biases and making the calculation of sampling error difficult. Finally, if the administrative data are used as a frame in addition to or in place of another one obtained from data collection, it may not be possible to analyze the issues of coverage and non-response. On the positive side, many administrative data sources are censuses, meaning that there will be no sampling error in the estimates obtained from them. Indicate the contribution to key estimates from administrative data. If used as a frame, report the imputation rate for item or complete non-response and explain how the imputation was performed. If the administrative data are simply summed to produce an estimate, include an estimate of the loss of precision due to imputation. If administrative data make up part of the estimate, the rest being accounted for by survey data, report the portion of the frame covered by the administrative data as well as the portion of the estimate. Produce a response rate combining both the administrative portion and the survey portion as explained in Trépanier et al. (2005).

7.4.3 Timeliness

This is a serious consideration for administrative data. It is common for this type of data to be unavailable until well after the reference period. In the case of using administrative data for a frame, it may be well out of date by the time it can be used. Additionally, if administrative data are integrated with survey data, it is important that the administrative data be as timely as the survey data, otherwise the entire process can be held up. Conversely, there are some cases where administrative systems are maintained in real time, making extraction of information from them much more timely than would be the use of a separate survey. Indicate the vintage of any administrative data used. Explain the assumptions that are made regarding the use of out-dated administrative data.

7.4.4 Coherence

This is another significant consideration with administrative data. This type of data is typically captured for another purpose and, as a result, will not necessarily mesh with already-defined concepts that might exist on other statistical holdings. This can be true in the case of concepts and definitions, and even in the sense of coverage and sample design. Administrative data might cover only a portion of the target population, making it problematic to use, or a sampling strategy may have been employed making the calculation of survey weights difficult to perform. There are cases where survey designers should have input into the design of the administrative systems, which can greatly increase the coherence of the data. List any exclusion that may complicate comparisons with other data. Indicators may include a measure of the target population not covered.

References

Babiyak, C. 2007. "Challenges in Collecting Police-Reported Crime Data." *ICES-III, Proceedings of the Third International Conference on Establishment Surveys, Survey Methods for Businesses, Farms, and Institutions*. Montreal, Quebec. June 18-21, 2007. p. 959-966.

Brackstone, G.J. 1987. "Issues in the use of administrative records for statistical purposes." *Survey Methodology*. Vol. 13. p. 29–43.

Brion, P.H. 2007. "Redesigning French Structural Business Statistics, Using More Administrative Data." *ICES-III, Proceedings of the Third International Conference on Establishment Surveys, Survey Methods for Businesses, Farms, and Institutions*. Montreal, Quebec. June 18-21, 2007.

Cox, L.H. and R.F. Boruch. 1988. "Record linkage, privacy and statistical policy." *Journal of Official Statistics*. Vol. 4, no. 1. p. 3–16.

Haziza, D., G. Kuromi, J. Bérubé. 2007. "Sampling and Estimation in the Presence of Tax Data in Business Surveys at Statistics Canada." *ICES-III, Proceedings of the Third International Conference on Establishment Surveys, Survey Methods for Businesses, Farms, and Institutions*. Montreal, Quebec. June 18-21, 2007.

Kovacevic, M. 1999. "Record linkage and statistical matching – they aren't the same!" *SSC Liaison*. Vol. 13, no. 3. p. 24–29.

Lavallée, P. 2000. "Combining Survey And Administrative Data: Discussion Paper." *ICES-II, Proceedings of the Second International Conference on Establishment Surveys, Survey Methods for Businesses, Farms, and Institutions*. Buffalo, New York. June 17-21, 2000. p. 841-844.

Lavallée, P. 2005. "Quality Indicators when Combining Survey Data and Administrative Data." *Proceedings of the XXII International Methodology Symposium*. Statistics Canada. Ottawa, Ontario. October 25-28, 2005.

McKenzie, R. 2007. "A Statistical Architecture for Economic Statistics." *ICES-III, Proceedings of the Third International Conference on Establishment Surveys, Survey Methods for Businesses, Farms, and Institutions*. Montreal, Quebec. June 18-21, 2007.

Michaud, S., D. Dolson, D. Adams, and M. Renaud. 1995. "Combining administrative and survey data to reduce respondent burden in longitudinal surveys." *Proceedings of the Section on Survey Research Methods*. American Statistical Association. p. 11–20.

Penneck, S. 2007. "The Future of Using Administrative Data Sources for Statistical Purposes." *ICES-III, Proceedings of the Third International Conference on Establishment Surveys, Survey Methods for Businesses, Farms, and Institutions*. Montreal, Quebec. June 18-21, 2007.

Statistics Canada. 1998. "Policy on Informing Survey Respondents." *Statistics Canada Policy Manual*. Section 1.1. Last updated March 4, 2009.
http://icn-rci.statcan.ca/10/10c/10c_001_e.htm.

Statistics Canada. 2005. *The Statistics Act*. Ottawa, Ontario.

Statistics Canada. 2008. "Policy on Record Linkage." *Statistics Canada Policy Manual*. Section 4.1. Last updated March 4, 2009.
http://icn-rci.statcan.ca/10/10c/10c_025_e.htm.

Trépanier, J., C. Julien, and J. Kovar. 2005. "Reporting Response Rates when Survey and Administrative Data are Combined." *Proceedings of the Federal Committee on Statistical Methodology Research Conference*. Arlington, Virginia. November 14-16, 2005.

Wallgren, A. and B. Wallgren. 2007. *Register-based Statistics: Administrative Data for Statistical Purposes*. New York. John Wiley and Sons, 258 p.

8 Response and non-response

8.1 Scope and purpose

Despite the best efforts of survey managers and operations staff to maximize response, most, if not all, surveys must deal with the problem of non-response.

Response refers here to all data obtained either directly from respondents or from administrative data. This broad definition of response is necessary to reflect the increased use of different collection strategies in the same survey, a practice that has become more and more commonplace. Moreover, as with survey data, administrative data is not exempt from non-response, whether it be partial or total. This non-response is sometimes the result of lateness in obtaining all the administrative data.

For a unit to be classified as responding, the degree of item response or partial response (where an accurate response is obtained for only some of the data items required from a respondent) must meet a minimum threshold level below which it is considered that there is unit non-response. In that case, the sampled person, household, business, institution, farm or other unit is classified as not having responded at all.

Classic non-response mechanisms are as follows: uniform non-response (or response missing entirely at random) where the response probability is completely independent of the units and the measurement process, and is constant over the entire population; non-response depending on an auxiliary variable [or response missing at random] where the response mechanism depends on certain auxiliary data or variables available for all units measured and non-response depending on the variable of interest [or response not missing at random] where the response probability depends on the variable of interest.

Non-response can have two effects on data: first, it introduces a bias in estimates when non-respondents differ from respondents in the characteristics measured; second, it contributes to an increase in the total variance of estimates since the sample size observed is reduced from that originally sought.

8.2 Principles

The degree to which efforts are made to get a response from a non-respondent depends on budget, time and staff constraints, impact on overall quality and the risk of non-response bias. If non-response persists, there are several approaches to reduce the effect of the non-response. Decisions on the degree of research to be undertaken to develop non-response adjustment techniques are subject to the constraints mentioned above.

In telephone or personal interviews, or during follow-up, try to collect as much basic information on the respondent as possible to avoid making adjustments based on assumptions a little later.

In dealing with non-response, take advantage of available auxiliary information as much as possible.

An effective respondent relations program, a well-designed questionnaire, the use of active management to ensure regular follow-up on collection operations and adaptive data collection (Laflamme, 2008) are essential elements in optimizing response.

8.3 Guidelines

8.3.1 Setting an anticipated response rate

One point to consider in determining sample size and managing collection is setting an anticipated response rate. One way of doing this is to use the results of previous survey cycles, a test run or similar surveys.

8.3.2 Reducing non-response

Ensure an acceptable level of quality in all survey planning and implementation steps to obtain a good response rate. To do this, keep the following factors in mind:

- When designing the survey: Previous experience with similar surveys, the total budget and allocation of the budget to various operations;
- The survey frame's quality (in terms of population coverage and the facility of establishing contact with respondents), the population observed and the sampling method;
- The data collection method (for example, by mail, personal interview or computer-assisted telephone interview, by electronic data reporting (EDR), the Internet or a combination of methods), the time of year and the length of the collection period;
- The communication strategy to be used to inform respondents of the importance of the survey and to maintain a relationship with respondents;
- The use and effectiveness of respondent incentives;
- The response burden imposed (length of interview, difficulty of subject matter, timing and interview periodicity); the subject's nature and sensitivity, questionnaire length and complexity; questionnaire language and respondents' cultural backgrounds;
- Collection staff's prior experience and skills in interpersonal relationships; their workload; factors related to the interviewers themselves, such as training; and potential staff turnover;
- The effectiveness and scope of follow-up methodology and expected difficulties in tracing respondents who have moved.

Institute adaptive collection allowing the collection strategy to evolve over time. This requires instituting active management for regular follow-up on collection operations and adaptive data collection in the four collection phases: before initial contact, after some trials, in the middle of the collection period and towards the end of collection.

8.3.3 Putting follow-up of non-respondents procedures in place during collection

Follow up on non-respondents (all or a subsample of them). Following up on non-respondents increases response rates and can help ascertain whether respondents and non-respondents are similar in the characteristics measured. The survey strategy should take non-response into account immediately by adopting a two-phase selection perspective.

Prioritize follow-up activities. For example, in business surveys, follow up on large or influential units first, possibly at the risk of missing the smallest units. Likewise, give high priority to non-responding units in known domains with a high potential for non-response bias. A score function can be used to prioritize the follow-up.

Follow-up is particularly important in the case of longitudinal surveys, in which the sample is subject to increasing attrition (and possibly bias) due to non-response on each survey occasion. In this case, high-quality tracing must be facilitated; obtain additional contact information for the units sampled in each survey cycle; provide a "Change of address" card and ask the sampled unit to advise the Bureau if it moves between survey cycles. This will help obtain up-to-date contact information. In addition, administrative data, city and telephone directories and many other sources, including local knowledge, are valuable to the tracing staff.

8.3.4 Assessing potential non-response bias

There are various approaches for determining whether there are differences between respondents and non-respondents and evaluating potential non-response bias: specific follow-up of units, follow-up of non-respondents and analysis of known characteristics of respondents and non-respondents. Information on non-respondents might come from previous information waves (in the case of longitudinal surveys or with rotation groups), or by using external data sources (e.g. administrative data or paradata files).

8.3.5 Determining the response mechanism

Analyzing respondents' and non-respondents' characteristics also helps establish a non-response model to reduce non-response bias as much as possible and determine the best way to compensate for non-response.

For longitudinal surveys, the structure of non-response over time must be considered (Hedeker and Gibbons, 2006).

8.3.6 Deciding how to handle non-response

The main approaches to dealing with missing data are imputation and reweighting.

The approach should be chosen based on the kind of non-response (total or partial), the availability of auxiliary variables and the quality of the response model. In general, reweighting is used to deal with total non-response. Imputation is mainly used for partial non-response although it may be used to deal with total non-response if auxiliary data is available (repeated surveys, administrative data, etc.)

Reweighting is used to eliminate, or at least reduce, total non-response bias. Reweighting can be viewed from two angles: non-response model or calibration (Särndal, 2007).

For the non-response model approach, a model is developed to estimate unknown response probabilities. The survey weights are then adjusted inversely to the estimated response probabilities (Oh and Scheuren, 1983; Lynn, P., 1996). To protect somewhat against model insufficiency, it is suggested that homogeneous response groups be formed, i.e. that units with the same characteristics and the same propensity to respond be grouped together (Haziza and Beaumont, 2007). Several methods may be used to do this: decision tree algorithms such as CHAID in Knowledge Seeker (Kass, 1980; Angoss Software, 1995), logistical regression models, the score method, use of auxiliary data such as paradata (Beaumont, 2005; Eltinge, Yansaneh, 1997).

Systems developed at Statistics Canada are used to evaluate and measure the impacts of non-response and imputation: GENESIS (GENERalized Simulation System) quantifies the relative performance of imputation methods using simulation studies and SEVANI (Système pour l'Estimation de la VARiance due à la Non-response et à l'Imputation) calculates variance due to non-response (Beaumont, 2007). It should be noted that if non-response variance is high compared to sampling variance for a given region, it might be useful to reduce the desired sample size and devote more resources to non-response prevention in order to stay within the budget.

8.3.7 Evaluating and disseminating non-response rates

Follow the Standards and Guidelines for Reporting Non-response Rates (Statistics Canada, 2001d) to make it easier to compare surveys. These standards describe non-response rate reporting requirements as set out in the Policy on advising users of data quality and methodology for censuses or sample surveys based strictly on collecting data directly from respondents. Subjects discussed include weighted and non-weighted non-response rates, response rates in data collection and estimation, non-response rates for secondary or longitudinal surveys, non-response bias, survey operations monitoring, data collection method evaluation, survey frame coverage measures, longitudinal database creation, non-response case reporting and integrated metadata base reporting requirements.

If need be, refer to particular implementations of the standards based on the specific characteristics of surveys. For example, recent articles have discussed surveys using administrative data for some units and survey data for others (Trepanier et al. 2005), surveys where a mixed collection method is used for the same unit (Leon, 2007) and random dialing surveys (Marchand, 2008).

8.3.8 Identifying and analyzing reasons for non-response

Note reasons for non-response at collection time (e.g. refusal, non-contact, temporary absence, technical problem) since non-response bias levels may differ depending on the reason.

8.4 Quality indicators

Main quality element: accuracy

8.4.1 Evaluating response and non-response rates

Write a note on the response rate. The response rate may be calculated in different ways, with interpretations for different purposes. Refer to Standards and Guidelines for Reporting Non-response Rates (Statistics Canada, 2001d). Report weighted response rates to show their contribution to estimates and use non-weighted response rates to reflect participation rates in the survey population.

Report non-response rates broken down by different non-response types. This information can be used later when designing other surveys and is useful to data users who must interpret the data. The percentages of sampled units that refused to respond, were identified as out of scope, could not be contacted during the collection period and partially responded might also be of interest.

Specify whether the survey estimates are adjusted to compensate for non-response. If estimates are adjusted, a description of the adjustment procedure should be appended.

8.4.2 Evaluating non-response variance

Report non-response variance. To do so, use SEVANI when the total or average of a domain or resampling methods is estimated.

8.4.3 Examining bias

Study the non-response bias based on the collection method and type of non-response.

In the case of periodic surveys, conduct periodic non-response bias studies. The results of these studies should be included in the information reported to users in accordance with the policy.

If applicable, try to determine how successful the procedures are in correcting potential bias.

References

Angoss Software. 1995. *Knowledge SEEKER – User's Guide*. ANGOSS Software International Limited.

Beaumont, J.-F. and J. Bissonnette. 2007. « Variance Estimation Under Composite Imputation Using an Imputation Model.» Article presented at the Workshop on Calibration and Estimation in Surveys, Ottawa, 2007.

Beaumont, J.-F. 2005. "On the Use of Data Collection Process Information for the Treatment of Unit Non-response Through Weight Adjustment." *Survey Methodology*. Vol. 31, no. 2. p. 227-231.

Eltinge, J.L. and I.S. Yansaneh. 1997. "Diagnostics for formation of non-response adjustment cells, with an application to income non-response in the U.S. Consumer Expenditure Survey." *Survey Methodology*. Vol. 23, no. 1. p. 33-40.

Fuller, W.A. 1993. *Measurement Error Models*. New York: Wiley-Interscience. 440 p.

Hedeker, D. and R.D. Gibbons. 2006. *Longitudinal Data Analysis*. New York. Wiley-Interscience..360 p.

Groves, R.M., D.A. Dillman, J.L. Eltinge and R.J.A. Little, 2001. *Survey Non-response*. New York. Wiley-Interscience. 520 p.

Haziza, D. and J.-F. Beaumont. 2007. "On the construction of imputation classes in surveys." *International Statistical Review*. Vol. 75, no. 1. p. 25–43.

Kass, G.V. 1980. "An exploratory technique for investigating large quantities of categorical data." *Applied Statistics*. Vol. 29, no. 2. p. 119-127.

Lynn, P. 1996. Weighting for Non-response. *Proceedings from the Association for Survey Computing 1996 Survey and Statistical Computing Conference*. London, UK.

Laflamme, F. 2008. Using Paradata to Actively Manage Data Collection Survey Process. *Proceedings from the American Statistical Society 2008 Joint Statistical Methods Conference*. Denver, Colorado.

Leon, C. A. 2007. Reporting Response Rates in Characteristic Surveys. *Proceedings from the Statistical Society of Canada 2007 Conference*. St. John's, Newfoundland.

Marchand, I., R. Chepita, P. St-Cyr, D. Williams. 2008. "Coverage and Non-Response in a Random Digit Dialling Survey: The Experience of the General Social Survey's Cycle 21 (2007)." Article presented at Symposium, Statistics Canada, 2008.

Oh, H.L. and F.J. Scheuren. 1983, "Weighting Adjustment for unit non-response", in W.G. Madow, I. Olkin, and D.B. Rubin (eds). *Incomplete data in Sample Surveys, Vol. 2: Theory and Bibliographies*. New York. Academic Press. p. 143-184.

Särndal, C.-E. and S. Lundström. 2005. *Estimation in Surveys with Non-response*. New York. Wiley. 212 p..

Statistics Canada. 2000d. "Policy on Informing Users of Data Quality and Methodology." *Statistics Canada Policy Manual*. Section 2.3. Last updated March 4, 2009.
http://icn-rci.statcan.ca/10/10c/10c_010_e.htm.

Statistics Canada. 2001d. *Standards and Guidelines for Reporting Non-response Rates*. Statistics Canada Technical Report.

Trépanier, J., C. Julien and J. Kovar. 2005. "Reporting Response Rates when Survey and Administrative Data are Combined." *Federal Committee on Statistical Methodology Research Conference*.

9 Editing

9.1 Scope and purpose

Data editing is the application of checks to detect missing, invalid or inconsistent entries or to point to data records that are potentially in error. Some of these checks involve logical relationships that follow directly from the concepts and definitions. Others are more empirical in nature or are obtained as a result of the application of statistical tests or procedures (e.g., outlier analysis techniques). The checks may be based on data from previous collections of the same survey or from other sources.

Editing encompasses numerous activities at various steps of a survey. It includes interviewer field checks and computer-generated warnings at the time of data collection and capture. It also includes identification of units for follow-up and detailed checks of the micro data during and after data. Finally, it includes error localization for the purposes of imputation, and complex relationship verifications at a macro level for the purposes of data validation.

9.2 Principles

A data record that has been altered as a result of edits should be closer to the truth after those alterations than before. We design edits to detect and correct inconsistencies, and not to generate bias by imposing implicit models. When further editing has a negligible impact on the final survey estimates, it is called over-editing, and should be avoided.

Analysis of edit failure rates and the magnitude of changes due to edits provide information about the quality of the survey data, and can also suggest improvements to the survey vehicle.

9.3 Guidelines

9.3.1 Design

- Editing is well suited for identifying fatal errors (Granquist and Kovar, 1997) - since the process can be easily automated. Perform this activity as quickly and as expeditiously as possible. While some manual intervention may be necessary, generalized, reusable software is particularly useful for this purpose. The Banff system for edit and imputation (Statistics Canada, 2009) and CANCEIS – the Canadian Census Edit and Imputation System (Bankier et al., 1999) – are examples of such software. Some customized applications can also be developed based on other software that does not target only editing processes. Logiplus – the Statistics Canada system for managing decision logic tables – is an example of such software.
- Automation allows and tempts survey managers to increase the scope and volume of checks that can be performed. Minimize any such increases if they make little difference to the estimates from the survey. Instead of increasing the editing effort, redirect resources into activities with a higher pay-off (e.g., data analysis, response error analysis.)
- Identify extreme data values in a survey period or across survey periods (this exercise is known as the outlier detection process). The presence of such outlying data is a warning sign of potential errors. Use simple univariate detection methods (Hidiroglou and Berthelot, 1986) or more complex and graphical methods (de Waal, 2000).
- The impact of errors has been shown to be highly variable, particularly in surveys that collect numeric data; it is not uncommon for a few errors to be responsible for the majority of changes in the estimates. Consider editing in a selective manner to achieve potential efficiency gains (Granquist and Kovar, 1997) without detrimental impact on data quality. Priorities may be set according to type or severity of error, or according to the importance of the variable or the reporting unit.
- Hit rates of edits, which are the proportions of warning or query edits that point to true errors, have been shown to be poor, often as low as 20-30%. Develop edits that are efficient and monitor the efficiency on a regular basis.

- Edits cannot possibly detect small, systematic errors reported consistently in repeated surveys, errors that can lead to serious biases in the estimates. “Tightening” the edits is not the solution. Use other methods, such as traditional quality control methods, careful analysis and review of concepts and definitions, post-interview studies, data validation, and data confrontation with other data sources that might be available for some units, to detect such systematic errors.
- Limit the reliance on editing to fix problems after the fact, especially in the case of repeated surveys. The contribution of editing to error reduction is limited. While some editing is essential, reduce its scope and redirect its purpose. Assign a high priority to learning from the editing process. To reduce errors, focus on the earlier phases of data collection rather than cleaning up at the end. Practice error prevention rather than error correction. To this end, move the editing step to the early stages of the survey process, preferably while the respondent is still available, for example, through the use of computer-assisted telephone or personal or self-interview methods.
- In designing data collection processes, especially editing and coding, make sure that the procedures are applied to all units of study as consistently and in as error-free a manner as possible. Automation is desirable. Enable the staff or systems to refer difficult cases to a small number of knowledgeable experts. Centralize the processing in order to reduce costs and make it simpler to take advantage of available expert knowledge. Given that there can be unexpected results in the collected information, use processes that can be adapted to make appropriate changes if necessary from the point of view of efficiency.

9.3.2 Data collection and failed edit follow-up

- Editing can serve a useful purpose in tidying up some of the data, but its much more useful role derives from its ability to provide information about the survey process, either as quality measures for the current survey or to suggest improvements for future surveys. Consider editing to be an integral part of the data collection process in its role of gathering intelligence about the process. In this role, editing can be invaluable in sharpening definitions, improving the survey vehicle, evaluating the quality for the data, identifying nonsampling error sources, serving as the basis of future improvement of the whole survey process, and providing valuable information to improve other survey processes and other surveys (Granquist, Kovar and Nordbotten, 2006). To accomplish this goal, monitor the process and produce audit trails, diagnostics and performance measures, and use these to identify best practices.
- When conducting follow-ups, do not overestimate the respondents’ ability to correct errors. Their aggregations may be different, their memory limited, and their “pay-off” negligible. Limit respondent follow-up activity.
- For business surveys, put in place a strategy for selective follow-up. The use of a score function (Latouche and Berthelot, 1992) concentrates resources on the important sample units, the key variables, and the most severe errors.

9.3.3 Quality assurance

- Ensure that all edits are internally consistent (i.e., not self-contradictory).
- Keep in mind that the usefulness of editing is limited, and the process can in fact be counter-productive. Often, data changes based on edits are erroneously considered as data corrections. It can be argued that a point is reached during the editing process when just as many errors are introduced as are corrected through the process. Identify and respect this logical end of the process.
- Reapply edits to units to which corrections were made to ensure that no further errors were introduced directly or indirectly by the correction process.
- Do not underestimate the ability of the editing process to fit the reported data to implicit models imposed by the edits. There exists a real danger of creating spurious changes just to ensure that the data pass the edits. Control the process!
- The editing process is often very complex. When editing is under the Agency’s control, make available detailed and up-to-date procedures with appropriate training to all staff involved, and monitor the work itself. Consider using formal quality control procedures.

- Monitor the frequency of edit rejects, the number and type of corrections applied by stratum, collection mode, processing type, data item and language of the collection. This will help in evaluating the quality of the data and the efficiency of the editing function.

9.4 Quality indicators

Main quality elements: accuracy, timeliness

Measurement error is the error that occurs during the reporting process while processing error is the error that occurs when processing data. The latter includes errors in data capture, coding, editing and tabulation of the data as well as in the assignment of survey weights. Although it is not usually possible to calculate measurement error and processing error individually, an indicator of their combined importance is given by the edit failure rate. This is the number of units rejected by edit checks divided by the total number of units. Outputs should be accompanied by a definition of the two types of error and a description of the main sources of errors. This informs users of the mechanisms in place to minimize error by ensuring fine tuned data collection, capture and processing. Measurement and processing error have impacts on bias and variance.

Editing rates for key variables should be reported. They may be higher due to measurement error (for instance, because of poor question wording) or because of processing error (for instance, data capture errors).

The total contribution to key estimates from edited values should be reported. This is the extent to which the values of key estimates are informed by data that have been edited. This may give an indication of the effect of measurement error on key estimates. This indicator applies only for means and totals.

Data editing is crucial to ensuring the accuracy and consistency of data. However it can become a costly and time consuming endeavour. It is likely the single most expensive activity of a sample survey or census cycle. When painstaking, often manual editing has a negligible impact on the final estimates, it is called over-editing. Not only is the practice of over-editing costly in terms of finances, timeliness and increased response burden, but it can also lead to severe biases resulting from fitting data to implicit models imposed by the edits.

References

- Bankier, M., M. Lachance and P. Poirier. 1999. "A generic implementation of the New Imputation Methodology." *Proceedings of the Survey Research Methods Section*. American Statistical Association. p. 548-553.
- De Waal, T., F. Van de Pol and R. Renssen. 2000. "Graphical macro editing: possibilities and pitfalls." *Proceedings of the Second International Conferences on Establishment Surveys*. Buffalo, New York.
- Granquist, L. and J.G. Kovar. 1997. "Editing of survey data: how much is enough?" *Survey Measurement and Process Quality*. Lyberg et al (eds.). New York. Wiley. p. 415-435.
- Granquist, L., J. Kovar and S. Nordbotten. 2006. "Improving Surveys: Where Does Editing Fit In?" *Statistical Data Editing*. Vol. 3: Impact on Data Quality, Chapter 4: "Looking forward". Conference of European Statisticians. United Nations Statistical Commission and United Nations Economic Commission for Europe.
- Hidiroglou, M. A. and J.-M. Berthelot. 1986. "Statistical editing and imputation for periodic business surveys." *Survey Methodology*. Vol. 12. p. 73-83.
- Latouche, M. and J.-M. Berthelot. 1992. "Use of a score function to prioritize and limit recontacts in editing business surveys." *Journal of Official Statistics*. Vol. 8. p. 389-400.
- Statistics Canada. 2009. *Functional Description of the Banff System for Edit and Imputation*. Statistics Canada Technical Report.

10 Imputation

10.1 Scope and purpose

Imputation is the process used to assign replacement values for missing, invalid or inconsistent data that have failed edits. This occurs after follow-up with respondents (if possible) and manual review and correction of questionnaires (if applicable). Imputation is typically used to treat item non-response and, occasionally, unit non-response. Unit non-response occurs when no usable information is collected for a given record while item non-response occurs when some but not all the desired information is collected. After imputation, the survey data file should normally only contain plausible and internally consistent data records that can then be used for estimation of the population quantities of interest.

10.2 Principles

Under the Fellegi-Holt principle (Fellegi and Holt, 1976), the fields to be imputed are determined by making changes to the minimum number of responded values so as to ensure that the completed record passes all of the edits. The determination of the fields to be imputed can be done before imputation or simultaneously with imputation.

Imputation is performed by those with full access to the microdata and thus in possession of auxiliary information known for both units with and without fields to be imputed. Auxiliary information can be used to predict missing values using a regression model; to find “close” donors to impute recipients; or to build imputation classes (e.g., Haziza and Beaumont, 2007). It can also be used directly as substitute values for the unknown missing values.

The basic principle underlying imputation is to use available auxiliary information in order to approximate as accurately as possible the unknown missing values and thus to produce quality estimates of population characteristics. Therefore, the application of this principle should normally lead to a reduction in both the bias and variance caused by not having observed all the desired values.

Good imputation processes are automated, objective and reproducible, make an efficient use of the available auxiliary information, have an audit trail for evaluation purposes and ensure that imputed records are internally consistent.

10.3 Guidelines

10.3.1 Auxiliary variables

- The choice of auxiliary variables used for imputation, also called matching variables for donor imputation, should mainly be based on the strength of their association with the variables to be imputed. To this end, consider using modelling techniques and consult subject matter experts to obtain information about variables.
- Consider using different sources of data (e.g. current survey data, historical data, administrative data, paradata, etc.) for variables that can be used as auxiliary variables for imputing the missing values. Study the quality and appropriateness of these available variables to determine which ones to ultimately use as auxiliary variables.
- Evaluate the type of non-response. That is, try to determine which auxiliary variables can explain the non-response mechanism(s) in order to use them to enrich the imputation method. Include such auxiliary variables in the imputation method, especially if they are also associated with the variables to be imputed.

- Take into account the type of characteristics to be estimated (such as level vs. change, high-level aggregates vs. small domains, and cross-sectional vs. longitudinal) when choosing auxiliary variables and developing an imputation strategy so as to preserve relationships of interest; e.g. use historical auxiliary information if you are interested in changes or use domain information (if available) if you are interested in domain estimation.

10.3.2 Imputation methods and their implementation

- Imputation methods can be classified as either deterministic or stochastic, depending on whether or not there is some degree of randomness in the imputation process (Kalton and Kasprzyk, 1986; Kovar and Whitridge, 1995). Deterministic imputation methods include logical imputation, historical (e.g. carry-forward) imputation, mean imputation, ratio and regression imputation and nearest-neighbour imputation. These methods can be further divided into methods that rely solely on deducing the imputed value from data available for the non-respondent and other auxiliary data (logical and historical) and those that make use of the observed data from other responding units for the given survey. Observed data from responding units can be used directly by transferring data from a chosen donor record or by means of explicit parametric models (ratio and regression). Stochastic imputation methods include the random hot deck, nearest neighbour imputation where a random selection is made from several “closest” nearest neighbours, regression with random residuals, and any other deterministic method with random residuals added.
- A serious modelling effort should normally be done to choose appropriate auxiliary variables and an appropriate imputation model. (An imputation model is a set of assumptions about the variables requiring imputation.) Once such a model has been found, the imputation strategy should be determined as much as possible in agreement with this model. This should help in controlling the non-response bias and variance and may be needed for proper variance estimation.
- Try to force the imputed record to be internally consistent but resemble as closely as possible the failed edit record. This is achieved by imputing the minimum number of variables in some sense, thereby preserving as much respondent data as possible (Fellegi-Holt principle). The underlying assumption is that a respondent is more likely to make only one or two errors rather than several, although this is not always true in practice.
- In some surveys, it is necessary to use several different types of imputation methods depending on the availability of auxiliary information. This is usually achieved in an automated hierarchy of methods. Carefully develop and test the methods used at each level of the hierarchy and limit as much as possible the number of such levels. Similarly, when collapsing of imputation classes is required, carefully develop and test the imputation methods for each set of classes.
- When donor imputation is used, try to impute data for a record from as few donors as possible. Operationally, this may be interpreted as one donor per section of questionnaire, since it is virtually impossible to treat all variables at once for a large questionnaire. Also, try to limit the number of times a specific donor is used to impute recipients in order to control the variance of imputed estimators. Based on available donors, this may imply allowing equally good imputation actions an appropriate chance of being selected to avoid artificially inflating the size of certain groups in the population.
- For large surveys, it may be necessary to process variables sequentially in two or more passes, rather than in a single pass, so as to reduce computational costs. As well, there may be extensive response errors on a record. Either of these conditions can make it difficult to follow the guidelines exactly: there may be cases where more than one donor is required (per section of the questionnaire), and more than the minimum number of variables are imputed.

10.3.3 Impact on estimates

- Information on the imputation process should be retained on the post-imputation files and be available for proper evaluation of the impact of imputation on estimates as well as on variances. Such information includes variables indicating which values were imputed and by which method, variables used to indicate which donors were used to impute recipients and so on. Retain the unimputed and imputed values of the record's fields for evaluation purposes.
- Consider the degree and impact of imputation when analyzing data. Even when the degree of imputation is low, changes to individual records may have a significant impact; for example, when changes are made to large units or when large changes are made to a few units. In general, the greater the degree and impact of imputation, the more judicious the analyst needs to be in using the data. In such cases, analyses may be misleading if the imputed values are treated as observed values.
- The imputation methods used may not preserve relationships between variables and may have a significant impact on distributions of data. For example, not much may have changed at the aggregate level, but values in one domain could have moved systematically up, while values in another domain could have moved down by an offsetting amount. This may actually mean that the domain variable needs to be taken into account in the imputation strategy.
- Evaluate the degree and effects of imputation. The Generalized System for Imputation Simulation (GENESIS) is one possible tool for this purpose. It performs imputation in a simulation environment and can be used to assess the bias and variance of imputed estimators in specific settings.

10.3.4 Generalized systems

- There exist a number of generalized systems that implement a variety of algorithms, either for continuous or categorical data. They should be considered during the development of the imputation methodology. The systems are usually simple to use once the edits are specified, and they include algorithms to determine which fields to impute. They are well documented and retain audit trails to allow evaluation of the imputation process. Two systems currently available at Statistics Canada are the Generalized Edit and Imputation System (GEIS/BANFF) (Kovar et al, 1988; Statistics Canada, 2000a) for quantitative economic variables and the Canadian Census Edit and Imputation System (CANCEIS) (Bankier et al, 1999) for qualitative and quantitative variables.

10.3.5 Variance estimation

- Consider the use of techniques to adequately measure the sampling variance under imputation and to measure the added variance caused by non-response and imputation (Lee et al, 2002; Haziza, 2008; Beaumont and Rancourt, 2005). This information is required to satisfy Statistics Canada's Policy on Informing Users of Data Quality and Methodology (Statistics Canada, 2000d; see Appendix 2 where this Policy is reproduced). The System for the Estimation of Variance due to Non-response and Imputation (SEVANI), developed at Statistics Canada, can be used for this purpose.
- The final report and recommendations of the Committee on Quality Measures (Beaumont, Brisebois, Haziza, Lavallée, Mohl, Rancourt and Trépanier, 2008) contain additional guidelines for variance estimation in the presence of imputation that should be read and considered before implementing any new methodology or software.

10.3.6 Resources

- To obtain general training on imputation or greater detail on some specific issues, there are different resources. First, it is suggested to take the Statistics Canada course "0423: Non-response and Imputation: Theory and application". The Imputation Bulletin is also an interesting and useful source of information on the subject. Finally, external consultants, such as David Haziza and J.N.K. Rao, as well as a number of internal consultants, including the members of the Statistical Research and Innovation Division, the members of the Committee on Quality Measures and the members of the Committee on Practices in Imputation, are available to answer questions.

10.4 Quality indicators

Main quality elements: accuracy, timeliness, interpretability, coherence.

Estimates obtained after non-response has been observed and imputation has been used to deal with this non-response are usually not equivalent to the estimates that would have been obtained had all the desired values been observed without error. The difference between these two types of estimates is called the non-response error. The non-response bias and variance (i.e. the bias and variance caused by not having observed all the desired values) are two quantities associated with the non-response error that are usually of interest. These unknown quantities, for which we would ideally like to obtain an accurate measure, are related to the 'accuracy' aspect of quality.

In theory, the non-response bias is eliminated if the imputation strategy is based on a correctly-specified imputation model with good predictive power. Such an imputation model also leads to a reduction of the non-response variance. An imputation model is correctly specified if, given the chosen auxiliary variables, the assumptions underlying its first moments (usually the mean and variance) hold. It is predictive if the chosen auxiliary variables are well associated with the variables to be imputed. As pointed out in the above guidelines, variables used in the definition of the estimator and variables associated with the non-response mechanism should be considered as potential auxiliary variables. The objective of these guidelines is to ensure that, given the chosen auxiliary variables, the respondents and non-respondents are similar with respect to the measured variables.

It is difficult to measure the magnitude of the non-response bias but it is possible to derive indicators that are associated with it. Since the magnitude of the non-response bias depends on the adequacy of the imputation model, standard model validation techniques, which can be found in classical textbooks on regression, can be used to derive useful indicators. For instance, graphs of model residuals versus different auxiliary variables, including predicted values, can be used to detect possible model misspecifications. The residuals could also be used to derive different statistics. For logistic regression, the Hosmer-Lemeshow test statistic may be a useful indicator. These indicators may also be useful for giving an idea of how the non-response variance has been controlled, especially those giving information on the strength of the association between the auxiliary variables and the variables to be imputed.

In addition to the above model diagnostics, estimates of the non-response variance or estimates of the total variance may provide good measures of the increased variability due to non-response provided that the non-response bias can be assumed to be reasonably small. The total variance is the sampling variance to which a non-response component is added to reflect the additional uncertainty due to non-response. Many variance estimation methods that take non-response and imputation into account exist as well as some software. For instance, estimates of the non-response component or the total variance can be obtained using SEVANI.

Other indicators can be considered and are useful to give an indication of the degree of imputation but are more difficult to directly relate to the non-response bias and variance. The imputation rate by variable and by important domains is one of these indicators. For estimates of totals and means, another useful indicator is the contribution to key estimates that comes from imputed values. A large contribution from imputed values may be an indication that the non-response bias and/or variance are not small. Other indicators of the impact of imputation on final estimates can also be determined, which provide other information to gauge the reliability of estimates.

As stressed in the above discussion, a serious modelling effort should be made before determining any imputation strategy. This requires time and resources. In practice, an appropriate balance must thus be struck between the time taken to produce the imputed data file (timeliness) and the quality of the underlying imputation model so as to avoid unduly delaying the release of data. Whenever appropriate, the use of generalized systems for imputation may contribute to substantially reduce the processing time, especially the time needed for system development, and thus ensure that more time can be devoted to the choice of a good imputation strategy.

Finally, the imputation methodology used should be clearly described and provided to users along with some of the above indicators and measures. This ensures a better interpretability of the survey results. As far as it is possible and relevant, the use of similar imputation methodologies across surveys collecting similar information should be considered for coherence purposes.

References

Bankier, M., M. Lachance. and P. Poirier. 1999. "A generic implementation of the New Imputation Methodology." *Proceedings of the Survey Research Methods Section*. American Statistical Association, 548-553.

Beaumont, J.-F., F. Brisebois, D. Haziza, P. Lavallée, C. Mohl, E. Rancourt, and J. Trépanier. 2008. *Final report and recommendations: Variance estimation in the presence of imputation*, Committee on Quality Measures. Statistics Canada technical report.

Beaumont, J.-F., and E. Rancourt. 2005. "Variance estimation in the presence of imputation at Statistics Canada." Paper presented at the Statistics Canada's Advisory Committee on Statistical Methods, May 2005.

Fellegi, I.P. and D. Holt. 1976. "A systematic approach to automatic edit and imputation." *Journal of the American Statistical Association*. Vol. 71. p. 17-35.

Haziza, D., and J.-F. Beaumont. 2007. "On the construction of imputation classes in surveys." *International Statistical Review*. Vol. 75. p. 25-43.

Haziza, D. 2008. "Imputation and inference in the presence of missing data." In *Handbook of Statistics*. Vol. 29. Chapter 10: Sample Surveys: Theory, Methods and Inference. D. Pfeffermann and C.R. Rao (eds.). Elsevier BV (to appear).

Kalton, G. and D. Kasprzyk. 1986. "The treatment of missing survey data." *Survey Methodology*. Vol. 12. p. 1-16.

Kovar, J.G., and P. Whitridge. 1995. "Imputation of business survey data." In *Business Survey Methods*. B.G. Cox et al. (eds.) New York. Wiley. p. 403-423.

Kovar, J.G., J. MacMillan and P. Whitridge. 1988. *Overview and strategy for the Generalized Edit and Imputation System*. Statistics Canada, Methodology Branch Working Paper no. BSMD 88-007 E/F.

Lee, H., E. Rancourt and C.-E. Särndal. 2002. "Variance estimation from survey data under single imputation." In *Survey Non-response*. R.M. Groves et al. (eds.) New York. Wiley. p. 315-328.

Statistics Canada. 2000d. "Policy on Informing Users of Data Quality and Methodology." *Statistics Canada Policy Manual*. Section 2.3. (Reproduced in Appendix 2). Last updated March 4, 2009. http://icn-rci.statcan.ca/10/10c/10c_010_e.htm.

Statistics Canada 2000a. *Functional description of the Generalized Edit and Imputation System*. Statistics Canada Technical Report.

11 Weighting and estimation

11.1 Scope and purpose

A typical survey objective is to estimate descriptive population parameters, as well as analytical parameters, on the basis of a sample selected from a population of interest. Examples of parameters include simple descriptive statistics such as totals, means, ratios and percentiles. Examples of analytical parameters include regression coefficients, correlation coefficients and measures of income inequality.

In a probability-based survey, a design weight is associated with each sampled unit. The design weight can be interpreted as the number of typical units in the survey population that each sampled unit represents. Estimates can be calculated using the design weights or estimation weights obtained by adjusting the design weights. Common adjustments include those that account for non-response and that incorporate auxiliary information. See Statistics Canada (2003).

The precision of an estimate is an important aspect of quality. This aspect is measured using the estimated standard error (square root of the estimated variance). Improvements to this precision can be obtained by incorporating auxiliary data into the estimation process.

11.2 Principles

In a probability-based survey, all elements of the population have a known probability of being selected into the sample. These inclusion probabilities take into account aspects of the sample design such as stratification, clustering and multi-stage or multi-phase selection. The design weight is equal to the inverse of the inclusion probability in single-phase (one stage) sampling. It is the product of the inverse selection probabilities at each phase (stage) in a multi-phase (multi-stage) design.

If there is unit non-response, the observed sample is smaller in size than the original sample selected. To compensate for unit non-response re-weighting can be performed by adjusting the design weights. These adjustment factors should be based on each unit's probability of response, which can be estimated using models.

If auxiliary data are available, some improvement in the precision of estimates may be achieved. Incorporation of auxiliary data in the estimation process is known as calibration. Calibration consists of adjusting the weights such that estimates of the auxiliary variable(s) satisfy known totals (also referred to as control totals). Calibration includes well-known estimators such as the regression, the ratio and the raking-ratio estimators (Deville and Särndal, 1992). Desirable properties of calibration include:

- Coherent estimates between different sources;
- Potential improvements to the precision of estimates;
- Potential reduction of unit non-response error and coverage error.

Estimates are produced by summing the data multiplied by either the design or estimation weights. There are two types of errors associated with these estimates: sampling and non-sampling errors. The sampling error is the error caused by observing a sample instead of the whole population (Särndal et al., 1992). It is measured by the sampling variance, which depends on the sample design, and any auxiliary data that are used in the estimation procedure. Non-sampling errors include coverage (imperfect frame), measurement, processing and non-response errors.

An estimate of the sampling variance can be calculated using methods such as Taylor linearization or resampling methods such as the jackknife and the bootstrap. Regardless of which method is used, it has to incorporate sample design properties such as stratification, clustering and multi-stage or multi-phase selection, if applicable.

It is more difficult to measure nonsampling errors. They may require additional data that are typically not available. Examples include repeated measures to evaluate measurement errors and recontact of non-respondents to evaluate non-response bias.

11.3 Guidelines

11.3.1 Weighting

- A weight needs to be associated with each sampled unit. This weight can be the design weight or the estimation weight (e.g., calibration weight). If only the design weight is used, the resulting estimator is called the Horvitz-Thompson estimator. If auxiliary information is used by calibration, the resulting estimator is called the calibration estimator. The weight associated with this estimator is known as the estimation or calibration weight. An estimation weight should be used whenever the design weight has been adjusted for non-response or auxiliary data.
- As full response is unlikely, non-response adjustments should be used to minimize the non-response bias. Applying these adjustments within subsets of the population can minimize the non-response bias. It is assumed that the non-respondents behave similarly to the respondents within these subsets. These subsets are delineated using auxiliary information (Lundström and Särndal, 2005) or propensity models (Eltinge and Yansaneh, 1997).
- If there are auxiliary data that are correlated with the variable(s) of interest, then calibration must be considered. These auxiliary data must be at least available for the sampled units, and the corresponding population totals must be known. The resulting calibration estimator will usually have a smaller variance than the Horvitz-Thompson estimator. In addition, the weighted auxiliary data will add up to the population totals.
- Calibration weights can be very large, or even negative. If this occurs, methods exist to control the range of the weights. See Huang and Fuller (1978) or Deville and Särndal (1992).
- Composite estimation should be considered for periodic surveys with a large sample overlap between occasions. It is a calibration method that treats the data from previous occasions as auxiliary variables. For more details, see Gambino, Kennedy and Singh (2001).
- Two sets of weights can be associated with longitudinal surveys: longitudinal weights and cross sectional weights. Longitudinal weights refer to the population at the initial selection of the longitudinal sample. For longitudinal analysis, these weights should be adjusted to account for sample attrition. Cross sectional weights reflect the population at a point in time. These weights can be used to produce point estimates, or differences between time periods.
- If double (two-phase) sampling has occurred, weighting needs to reflect the design and any auxiliary data available for the population or the first phase sample.

11.3.2 Estimation

- The estimation process must use the estimation weights to compute descriptive and analytical statistics for domains of interest. The estimation weights are equivalent to the design weights if no adjustments have been made. Corresponding variance estimators must reflect the sample design, any adjustments to the design weights, imputation and the estimation method. Variances can be estimated using linearization methods or resampling methods (jackknife, balanced repeated replication, and bootstrap). For more details, see Wolter (2007).

- Small domains refer to subpopulations where there is not enough sample (or even no sample at all) to produce reliable estimates. It is therefore reasonable to incorporate the requirements of these domains at the sample design stage (Singh, Gambino and Mantel, 1994). If this is not possible at the design stage, or if the domains are only specified at a later stage, consider special estimation methods (small area estimators) at the estimation stage. These methods “borrow strength” from related areas (or domains) to minimize the mean squared error of the resulting estimator (Rao, 2003).
- Generalized estimation software should be used when appropriate (Estevao et al., 1995).

11.4 Quality indicators

Main quality element: accuracy

- The quality of a point estimate is usually described in terms of accuracy and precision. Accuracy represents how closely a measured value agrees, on average, with the true value. The accuracy of an estimator is gauged in terms of how close the average of its realized values is to the parameter of interest. This is measured by comparing its design expectation to the parameter, and the difference is called the bias. Precision, on the other hand, refers to how closely individual measurements agree with each other. Precision is usually measured by the sampling error: it is the error that results from observing a sample instead of the whole population. If an estimator is unbiased, its mean squared error is equal to its sampling variance.
- Unbiased estimators should be used if they exist and are efficient in terms of variance. Slightly biased estimators can be used if their efficiency, measured in terms of mean squared error, is smaller than the variance of corresponding unbiased estimators.
- The coefficient of variation is normally used to describe the precision of an estimate. It is defined as the standard error of the estimate divided by the true value of the parameter. An estimate with a given coefficient of variation is less precise than one with a smaller coefficient of variation. Because of potential division by zero, as well as interpretation problems, the use of coefficients of variation should be restricted to positive variables of interest. Otherwise, standard errors should be used.
- Estimators that incorporate auxiliary data assume that the models between the target variables and auxiliary data hold for all units in the population. In practice, however, it is difficult to determine whether model assumptions are valid. Estimates that use auxiliary data should be accompanied by a description of the model assumptions made and an assessment of the likely effect of making these assumptions on the quality of estimates.

References

Deville, J.-C. and C.E. Särndal. 1992. “Calibration estimators in survey sampling.” *Journal of the American Statistical Association*. Vol. 87. p. 376-382.

Eltinge, J.L. and I.S. Yansaneh. 1997. “Diagnostics for formation of non-response adjustment cells, with an application to income non-response in the U.S. Consumer Expenditure Survey.” *Survey Methodology*. Vol. 23. p. 33-40.

Estevao, V., M.A. Hidirolou and C.E. Särndal. 1995. “Methodological principles for a generalized estimation system at Statistics Canada.” *Journal of Official Statistics*. Vol. 11, p. 181-204.

Gambino, J., B. Kennedy and M.P. Singh. 2001. “Regression composite estimation for the Canadian Labour Force Survey: Evaluation and implementation.” *Survey Methodology*. Vol. 27, no. 1. p. 65-74.

Huang, E. T. and W.A. Fuller. 1978. “Nonnegative regression estimation for sample survey data.” *Proceedings of the Social Statistics Section, American Statistical Association*. p. 300-303.

Lundström, S. and C.-E. Särndal. 2005. *Estimation in Surveys with Non-response*. New York. John Wiley and Sons. 212 p.

Rao, J.N.K. 2003. *Small Area Estimation*. New York. John Wiley and Sons. 344 p.

Särndal, C.E., B. Swensson and J.H. Wretman. 1992. *Model Assisted Survey Sampling*. New York. Springer-Verlag. 694 p. Springer Series in Statistics.

Singh, M.P., J. Gambino. and H. Mantel 1994. "Issues and strategies for small area data." *Survey Methodology*. Vol. 20. p. 3-14.

Statistics Canada 2003. *Survey methods and practices*. Statistics Canada Catalogue no. 12-587-XPE, Ottawa, Ontario. 396 p.

Wolter, K. 2007. *Introduction to Variance Estimation*. 2nd ed. New York. Springer-Verlag.. 449 p.

12 Seasonal adjustment and trend-cycle estimation

The Seasonal Adjustment Guidelines for Statistics Canada's Methods and Standards Committee were written in March 2000. This is a proposal to update them following Statistics Canada's adoption of X-12-ARIMA (Findley et al., 1998) and to standardize them with those at the US Census Bureau (McDonal-Johnson et al., 2006a, 2006b) and EUROSTAT (Mazzi, 2008).

12.1 Scope and Purpose

12.1.1 Seasonal adjustment

A time series is a sequence of measurements of one variable observed through time. In most situations, the observations are dependent through time and it is this dependence in itself that is of interest. For the purpose of seasonal adjustment, the time series is assumed to be observed either monthly or quarterly and constituted of three distinct elements: the trend-cycle, the combined seasonal and calendar effects and the irregular. The objective of seasonal adjustment is to identify and estimate the combined seasonal and calendar effects, and to remove them from the time series. The resulting series is then called seasonally adjusted and consists of only the trend-cycle and irregular components.

Seasonal effects are the intra-year (monthly, quarterly) fluctuations which repeat more or less regularly from year to year. They result from composite effects of events related to the climate, institutional decisions or modes of operation which repeat with a certain regularity within the year. Calendar effects are related to the composition of the calendar. They include trading-day effects associated with the weekday composition of the month, moving holiday effects associated with non-fixed date holidays such as Easter, and other predictable events from the calendar. Trading-day effects are present when the level of activity varies with the days of the week. The Easter effect may be seen as the variation in level due to the displacement of a volume of activity from April to March when Easter falls in March instead of the usual April occurrence.

The seasonally adjusted series enables the assessment of the current trend-cycle direction with month-to-month or quarter-to-quarter comparisons.

The trend is the underlying long-term movement lasting many years. The cycle, also called business-cycle, is a quasi-periodic oscillation lasting for more than a year around the long-term trend. It is characterized by alternating periods of expansion and contraction. The trend and the cycle are difficult to estimate separately and thus are considered and analyzed as a whole as the trend-cycle.

The irregular component represents random variations that are unforeseeable movements related to events of all kinds and which cannot be attributed to the trend-cycle component, the seasonal component or the calendar effects.

12.1.2 Trend-cycle estimation

Seasonal adjustment of highly volatile series may not be enough to draw conclusion on the current trend-cycle direction. In those cases, further smoothing of the seasonally adjusted series is advisable to eliminate most of the irregular component. The resulting trend-cycle estimate is to be considered auxiliary information to the seasonally adjusted series.

12.1.3 X-12-ARIMA

The foundation of Statistics Canada's seasonal adjustment program is 1967's X-11 Variant of the Census Method II (Shiskin et al., 1967; Ladiray and Quenneville, 2001). In 1980, Statistics Canada incorporated the autoregressive integrated moving average (ARIMA) (Box and Jenkins, 1976) models to forecast and backcast the series before seasonal adjustment as well as several major modifications to improve the Census method. This new variant was called X-11-ARIMA (Dagum, 1980). Extending the series with ARIMA forecasts gave, on average, smaller revisions to the seasonally adjusted series. In 1988, Statistics Canada released an improved version of X-11-ARIMA (Dagum, 1988) available for microcomputers.

In 1998, the US Census Bureau released X-12-ARIMA (Findley et al., 1998). X-12-ARIMA uses linear regression with ARIMA errors (regARIMA modeling) to estimate calendar effects as well as additive outliers, level shifts and other predefined regression variables. Additionally, X-12-ARIMA permits user-defined regression for unusual or nonstandard calendar effects and includes a variant of the TRAMO algorithm (Gomez and Maravall, 1996) for automatic regARIMA modeling. Other features are described in the user guide (US Census Bureau, 2008). X-12-ARIMA is available for several computing platforms, including its raw FORTRAN executable, C for UNIX and one for FAME. The X12 procedure (SAS Institute Inc, 2007) in SAS® also implements the most important options of the method. X-12-ARIMA is the recommended seasonal adjustment method at Statistics Canada as X-11-ARIMA is being phased-out and will no longer be supported.

12.2 Principles of seasonal adjustment

As seasonal adjustment aims to filter out the combined seasonal and calendar effects, it should be applied when those effects can be properly identified and estimated. When seasonal and/or calendar effects cannot be identified in a time series, the series is considered de facto seasonally adjusted. Seasonally adjusted series should have no residual seasonality and are generally smoother than the corresponding raw series.

As new data becomes available, the various time series components can better be estimated. This results in revised and more accurate estimates for past seasonally adjusted values which should be acknowledged. However, too frequent revisions may hinder the usefulness of the seasonally adjusted data. Seasonally adjustment options should be selected to minimize the amplitude of the revisions without affecting the overall quality of the adjustment and a revision strategy should be implemented to minimize the frequency of the revisions to the published data.

Inappropriate seasonal adjustment options may lead to misleading results. As such proper time and effort should be put into the analysis of the series and on the initial selection and maintenance of options. Similarly, as various sets of options could lead to various results, collection of series which measure the same economic activity should be treated as a whole. This usually entails the use of similar adjustment options by the area involved for coherence purposes.

12.3 Principles of trend-cycle estimation

As a complement to the seasonally adjusted series, trend-cycle estimates may indicate the direction of the short-term trend (within the current year). When new data points are added to the series, past trend-cycle estimates can be estimated more accurately and are therefore subject to revisions. Trend-cycle estimates are sensitive to the current phase of the business cycle (turning point, recession, recovery or expansion); thus, the reliability of the current trend-cycle estimates depends on the proximity of a turning point as well as on the amplitude of the cycle.

Trend-cycle estimates should be fully consistent with the published seasonally adjusted series. If the seasonally adjusted values are frozen in a database, the trend-cycle should be estimated from the seasonally adjusted series as it appears in the frozen database. Similarly, if the seasonally adjusted series underwent further adjustments such as aggregation, balancing or reconciliation, the trend-cycle should be estimated from the aggregated, balanced or reconciled series.

12.4 Seasonal adjustment guidelines

- Before seasonally adjusting a series for the first time, one should assess if the seasonality is identifiable and if it can properly be estimated.
- If a series is neither seasonal or does not have calendar effects, no treatments are applied and the series is deemed de facto seasonally adjusted.
- Seasonally adjusted series should not have residual seasonality, nor residual calendar effects.
- For proper identification and estimation of seasonal and calendar effects, it is recommended to use a span of 10 to 15 years of data. The minimum is five (5) years to properly estimate a seasonal pattern and seven (7) years for calendar effects such as trading days and moving holidays.
- It is recommended to use regARIMA modeling to calculate calendar adjustment factors and temporary adjustments such as those from known additive outliers or level shifts. The same regARIMA model should usually be used to extrapolate the series in order to reduce revisions in the seasonally adjusted series.
- There are various seasonal adjustment options; the most important ones revolve around the selection of the decomposition model, the specification of a regARIMA model and the seasonal and trend-cycle filter lengths. X-12-ARIMA's automatic selection processes may be used for a preliminary setting of those options. When time permits or when quality expectation is high, the automatic selection should be reviewed using alternatives statistics, expert prior knowledge and graphical analysis.
- Seasonal adjustment options for each series should be reviewed periodically to verify their continued applicability and appropriateness, and to increase accuracy. The already selected seasonal adjustment options should not be changed between scheduled reviews without a right justification.
- Although the main seasonal adjustment options should usually be fixed between reviews, the adjustment factors and the regARIMA model parameters should be concurrent – that is, recomputed using all available data points. Exceptions may apply when the most recent observations have been historically subjected to large revisions. In this case, year-ahead (forecasted) factors may be more appropriate.
- For an aggregate (or composite) series comprising several component series, seasonal adjustment can be done indirectly – the seasonally adjusted components are aggregated to form the seasonally adjusted composite series – or directly – the aggregate is adjusted independently. With this last method, there can be discrepancies between the aggregate series and the aggregated components after seasonal adjustment. When required, apply a raking or a reconciliation method to reconcile the direct seasonally adjusted aggregate series with its seasonally adjusted components, without modifying the unadjusted components if possible.
- For either the indirect and direct approach, the aggregate should not contain residual seasonality and should be relatively smooth. The raked direct approach is preferred when more importance is given to the aggregate series than its components or when the components have very similar observed seasonal components. The indirect adjustment is usually appropriate when the component series have very different seasonal patterns and many of the series can be seasonally adjusted individually.
- Forcing the annual totals of the seasonally adjusted data to be equal to those of the original (or calendar adjusted original) series is rarely theoretically justified but could be used when there is a need for consistency with external benchmarks such as in the System of National Accounts or in reconciling an aggregate and its components.

12.4.1 Revisions to seasonally adjusted data

- Revisions to the seasonally adjusted data are to be published according to an officially stated revision policy and in alignment with the release calendar of the unadjusted data.
- When a concurrent seasonal factor is used, it is not necessary to revise the seasonally adjusted estimates more than one period back. Exceptions may apply when preliminary observations are used: it is recommended to revise the seasonal factors whenever the original figures are revised. On an annual basis, revise the seasonally adjusted values for the last three years when the first month (quarter) of the next year becomes available. When seasonally adjusted values are obtained with year-ahead (forecasted) seasonal adjustment factors, the annual revision applies to the last four years.

12.4.2 Trend-cycle estimation

- Apply the trend-cycle estimation method to the published seasonally adjusted series to ensure that the trend-line is centered on the seasonally adjusted series. The Dagum (1996) method or an appropriate adaptation/variant of it is the recommended methods for trend-cycle estimation.
- Inform the users that the last few trend-cycle estimates (and especially the very last estimate) are subject to revisions when one more data point is added. This higher variability associated with the estimates around the end can be indicated, for example, by a dashed line on the trend graph or by publishing an information note with the data.
- Revise trend estimates as far back as the seasonally adjusted estimates were revised, and in a typical monthly series add 3 more months (2 for quarterly) during the year, and 6 months (2 for quarterly) at the annual review.

12.4.3 Data presentation and data access

- Month-to-month (or quarter-to-quarter) growth rates and changes should be computed on seasonally adjusted data and should be used with caution if the time series has high volatility. Year on year same-month comparisons should be computed on calendar adjusted data, or, in absence of calendar effects, on raw data.
- Users should be given access to the full historical raw series, the seasonally adjusted and, upon request, to the seasonal adjustment options.

12.4.4 Implementation

Assistance with the interpretation and implementation of these guidelines can be obtained from the Time Series Research and Analysis Centre (TSRAC), Business Survey Methods Division.

12.5 Quality indicators

The following indicators may be used to assess if a series is seasonal:

- A simple time plots and a year-over-year graph to inspect the series for seasonal patterns and to visually find other perturbations.
- Various statistics such as the two Fisher tests for stable and moving seasonality and spectrum graphs described in (Ladiray and Quenneville, 2001; Findley et al., 1998).

Residual seasonality can be tested as above on the seasonally adjusted data or with other tests. (Ladiray and Quenneville, 2001)

Statistics to assess the significance of the estimated regARIMA components and the overall quality of the fitted model are described in many textbooks on the subject.

Simple summary statistics on the historical revisions of both the level and the period-to-period change in the seasonally adjusted data may be used to quantify the revisions. Revisions in the seasonal adjustment context usually improve accuracy because they come from using observations that were not initially available.

References

- Box, G.E.P. and G.M. Jenkins. 1976. *Time Series Analysis, Forecasting and Control*. San Francisco. Holden Day. 575 p.
- Dagum, E.B. 1980. *The X11ARIMA Seasonal Adjustment Method*. Statistics Canada Catalogue no. 12-564E.
- Dagum, E.B. 1988. *The X11ARIMA/88 Seasonal Adjustment Method - Foundations and User's Manual*. Time Series Research and Analysis Division. Statistics Canada Technical Report.
- Dagum, E.B. 1996. "A New Method to Reduce Unwanted Ripples and Revisions in Trend-Cycle Estimates from X-11-ARIMA." *Survey Methodology*. Vol. 22. p. 77-83.
- Findley, D.F., B.C. Monsell, W.R. Bell, M.C. Otto, and B.C. Chen. 1998. "New Capabilities of the X-12-ARIMA Seasonal Adjustment Program." With Discussion. *Journal of Business and Economic Statistics*. Vol. 16. p. 127-77.
- Gomez, V. and A. Maravall. 1996, Programs TRAMO (Time series Regression with Arima noise, Missing observations, and Outliers) and SEATS (Signal Extraction in Arima Time Series). Instructions for the User. Working Paper 9628, Research Department, Banco de España. <http://www.bde.es/servicio/software/tramo/summprogs.pdf>.
- Ladiray, D. and B. Quenneville. 2001. *Seasonal Adjustment with the X-11 Method*. New York: Springer-Verlag. 256 p. Lecture Notes in Statistics. Vol. 158.
- Mazzi, G.L. 2008. *ESS Guidelines on Seasonal Adjustment*. http://epp.eurostat.ec.europa.eu/pls/portal/docs/PAGE/PGP_RESEARCH/PGE_RESEARCH_04/ESS%20GUIDELINES%20ON%20SA.PDF.
- McDonal-Johnson, K.M., B. Monsell, R. Frscina, and R. Feldpaush. 2006. *Seasonal Adjustment Diagnostics, Census Bureau Guideline*. Version 1.0. Washington. US Census Bureau.
- McDonal-Johnson, K.M., B. Monsell, R. Frscina, and R. Feldpaush. 2006. "Supporting Document A, Seasonal Adjustment Diagnostics Checklists." *Census Bureau Guideline*. Version 1.0. Washington. US Census Bureau.
- SAS Institute Inc. 2007. *SAS/ETS® 9.2 User's Guide*. Cary, NC. SAS Institute Inc.
- Shiskin, J., A.H. Young and J.C. Musgrave. 1967. *The X-11 Variant of the Census Method II Seasonal Adjustment*. Technical Paper No. 15, Bureau of the Census, U.S. Department of Commerce.
- US Census Bureau. 2008. *X-12-ARIMA Reference Manual*. Statistical Research Division, Census Bureau.

13 Benchmarking and related techniques

13.1 Scope and purpose

Statistical programs often have two sources of data measuring the same target variable: A more frequent measurement with an emphasis on an accurate estimation of the period to period movement and a less frequent measurement with an emphasis on an accurate estimation of the level. Without loss of generality, the more frequent series will from hereon be referred to as a sub-annual series, whereas the less frequent series will be used as benchmark and considered to be an annual series.

Benchmarking refers to techniques used to ensure coherence between time series data of the same target variable measured at different frequencies, for example, sub-annually and annually. Benchmarking consists of imposing the level of the benchmark series while minimizing the revisions of the observed movement in the sub-annual series as much as possible. Consequently, the growth rates in the benchmarked series are coherent with those from the benchmarks. In certain situations, benchmarking can improve the accuracy and timeliness of statistical output.

Non-binding benchmarking, interpolation, temporal distribution, calendarization, linkage and reconciliation are related techniques which are based on similar methodological principles and guidelines as those of benchmarking. Non-binding benchmarking is used when the benchmark series can also be revised. Interpolation is the estimation of intermediate terms between known values and can also be used to benchmark stock series. Temporal Distribution is the disaggregation of the benchmark series into more frequent observations. Calendarization is a special case of temporal distribution. Linkage is used to join different time series segments into a consistent single time series. Reconciliation is used to impose cross-sectional additive constraints among the components of a system of time series. More details on those techniques can be found in Dagum and Cholette (2006).

Benchmarking in the context of time series should not be confused with weight adjustments that can be applied at the estimation stage for calibration purposes.

13.2 Principles

One should make sure that there are as few conceptual, methodological and operational discrepancies as possible between the two data sources at the design stage. Discrepancies between the series should be thoroughly investigated and understood, after which an informed decision on whether to publish the series as is or to benchmark the series to ensure full numerical coherence can be made. In the former case, the discrepancies should be explained to users as per the Policy on Informing Users of Data Quality and Methodology (Statistics Canada, 1998).

In the case where the data source's designs are compatible or when external constraints impose the need of full numerical coherence, benchmarking methods can and – from a statistical point of view – should be used. In typical situations, the sub-annual series is implicitly assumed to be less reliable than the annual data. By its nature, the benchmarking process will cause various sets of revisions to the sub-annual data and thus, a willingness to revise is necessary.

All the related techniques are based on similar principles: The underlying assumptions should first be understood and the applicability of the methods should be verified – most frequently by a thorough analysis of the data before and after the techniques are used.

13.3 Guidelines

- Before considering benchmarking, investigate, document and quantify the discrepancies between the two sources of data. These differences should be minimized as much as possible at the design stage.
- Before considering benchmarking, examine differences in the microdata for common sample units, if any. Applied corrections, if made, must respect the time series nature of the data. For the sub-annual data, corrections should attempt to improve the accuracy of the period to period movement; for the annual data, corrections must consider both the accuracy of the level and the accuracy of the year to year movement.
- Be aware that the design of the annual series may not be compatible with the goals of benchmarking. For benchmarking, the annual survey has to provide both an accurate measurement of the annual level and an accurate measurement of the annual change since they will be imposed on the benchmarked series.
- Do not benchmark when annual values are less reliable than the annual sums from the sub-annual series. In this case, imposing the annual benchmarks will essentially produce less reliable benchmarked series.
- When the data sources are designed differently, consider benchmarking only if strong external constraints impose the need of full numerical consistency. Be aware that the resulting coherence may come at a cost of reduced accuracy.
- Benchmarking will result in revisions of the sub-annual data. Only consider benchmarking when the gain in consistency strongly reduces confusion among the users or the improved accuracy gained from a high quality annual series outweighs the burden of repeated revisions.
- Benchmark in the context of seasonal adjustment when there are unwanted discrepancies between the yearly sums of the raw and the corresponding yearly sums of the seasonally adjusted series. When required, seasonally adjusted series may be benchmarked to yearly totals derived from the raw series.
- Use an appropriate benchmarking method such as the regression-based techniques described in Dagum and Cholette (2006) or one of the various enhanced Denton methods described in the Quarterly National Accounts Manual from the International Monetary Fund (Bloem et al, 2001). Avoid simple pro-rating techniques as they introduce discontinuities between the years (known as the step-problem).
- Understand the underlying assumptions when the most recent observations of the sub-annual series are missing a corresponding annual value -- either because the year is incomplete or the annual data is not yet available. Benchmarking techniques will use either an implicit or explicit projection of the next annual value; projections can be based on short-term or long-term historical data or on external considerations.
- When implementing benchmarking or related techniques, consider the use of generalized software. This minimizes programming error and reduces development cost and time. Support from Methodology and Informatics is available within Statistics Canada, especially on the following software: in-house SAS Proc Benchmarking – for benchmarking, temporal distribution and linkage; in-house SAS Proc TSraking – for reconciliation; SAS Proc Expand – for interpolation; the US Bureau of the Census X-12-ARIMA program, SAS Proc X12, or SAS Proc ARIMA – for statistical inference methods on time series.
- Assistance with the interpretation and implementation of these guidelines can be obtained from the Time Series Research and Analysis Centre (TSRAC), Business Survey Methods Division.

13.4 Quality indicators

The following indicators may be used to document and quantify the discrepancies:

- Descriptive statements of conceptual aspects: reporting periods of the annual source; definition of the variables measured and of the target population, etc...
- Descriptive statements of operational aspects: sampling frame, collection process, etc...
- Descriptive statements of methodological aspects: sampling, use and source of administrative data, etc...
- When appropriate, quantitative descriptions of the discrepancies: counts and weighted counts by reporting periods to estimate the impact of the non-calendarization of annual data, differences between administrative data, sampling errors of the two annual estimates and corresponding annual changes, etc...

For more details and a case study, see Yung et al (2008).

Applying the benchmarking techniques and thoroughly analysing the results may also provide a good indicator of the appropriateness of the method. Annual discrepancies, revision to the series through the benchmarked-to-indicator (BI) ratios and revision to growth rates may all be studied with graphs or summary statistics.

References

Bloem, A. M., R. J. Dippelsman, and N. Ø. Mæhel. 2001. *Quarterly National Accounts Manual, Concepts, Data Sources and Compilation*. International Monetary Fund, Washington DC.

Dagum, E.B. and P.A. Cholette. 2006. *Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series*. New York. Springer. 410 p. Lecture Notes in Statistics #186.

Statistics Canada. 1998. "Policy on Informing Survey Respondents. *Statistics Canada Policy Manual*, Section 1.1. Last updated March 4, 2009. http://icn-rci.statcan.ca/10/10c/10c_001_e.htm.

Yung, W., B. Brisebois, C. Tardif, G. Kuromi, and C. Rondeau. 2008, *Should Sub-Annual Surveys be Benchmarked to their Annual Counterparts? A Case Study of Manufacturing Surveys*, Working Paper BSMD-2008-001, Statistics Canada. Ottawa, Ontario.

14 Data quality evaluation

14.1 Scope and purpose

Sound data quality practices are built into all survey steps, as described in the other chapters of this document. A data quality evaluation is a process to determine how well final products meet the original objectives of the statistical activity, in particular in terms of the reliability from an accuracy, timeliness and coherence point of view. It allows users to better interpret survey results and the Agency to improve the quality of its surveys.

There are two broad methods of evaluating data quality:

- Certification or validation: data are analysed before official release with a view to avoiding gross errors and eliminating poor quality data; comparisons with external or auxiliary sources are often the tool of choice at this stage;
- Studies of the sources of error: these generally provide quantitative information on the specific sources of errors in the data.

The data quality evaluation draws from the quality indicators produced in all survey steps. Methods for evaluating quality at each of those steps are given through out this document. While a single one-dimensional index of quality is often thought impossible to compile, the various indicators can be summarized and compared in terms of their relative importance and consequences.

14.2 Principles

Data quality evaluations should be conducted to determine to what extent the statistical information is relevant and representative.. However, users are rarely able to independently evaluate the quality of data produced by a statistical agency. It is therefore up to each agency to evaluate data quality and quickly provide users with the results in a usable form.

Data quality evaluations should be conducted to determine the extent to which errors can be associated with certain stages of the survey process. Such evaluations can be used to improve the quality of the next iteration of the survey, as well as other similar surveys.

Data quality evaluations at Statistics Canada must be designed to meet the mandatory and minimum requirements of the Policy on Informing Users of Data Quality and Methodology (Statistics Canada, 2000). The minimum requirements include measuring or evaluating coverage errors, response or imputation rate and (if dealing with a sample survey) measuring sampling errors of key characteristics.

14.3 Guidelines

14.3.1 Design

- Determine the extent of data quality evaluation required for a program or a product. The factors to be considered are: data uses and users; risk of errors and impact of errors on data use; quality variation over time; cost of the evaluation in relation to the total cost of the program; improving quality; increasing efficiency and productivity; usefulness of measures to users and ease of interpretation; and whether the survey has been, will be or will not be repeated.
- Make planning of data quality evaluations part of the overall survey design, as the information needed for such evaluations must often be collected during the survey process. Data quality reports should be included in the dissemination schedule for the survey.
- Data quality evaluation results should be valid and timely enough to improve released data. When this is not possible, evaluation results should at least be timely enough to help users to analyse the data and survey takers to improve the design of the next iteration of the survey or similar surveys.

14.3.2 Execution

- Provide a quality evaluation based on expert opinion or subjective analysis whenever data quality evaluations will not yield quantitative measurements because of the nature of the product, the user, time constraints, cost or technical feasibility.
- In the case of repeated surveys and statistical activities, it may not be necessary, or even possible, to consistently produce detailed quality evaluations. However, periodically review activity to ensure it meets its objectives – not just when problems arise.
- Involve users of evaluation results, whether they are associated with a statistical agency or not, in establishing the data quality evaluation program objectives. When circumstances permit, also involve them in the evaluation process.
- In order to perform all of these evaluations, the survey manager or the survey management team must have identified targets or standards which they want to attain.

14.3.3 Certification or validation

- Certification or validation of the statistical information should be conducted whenever appropriate or possible.
- Certification or validation should challenge rather than rationalize the data. It is recommended to involve analysts who did not take part in the production of the data.
- Check coherence in relation to external data sources such as other surveys, other iterations of the same survey or administrative data.
- Check internal coherence – for example: by calculating ratios that are known to be within certain limits (male-female ratios, average values of properties, etc).
- Analyse the largest units individually with regard to their contribution to overall estimates (generally applied to business surveys).
- Review and interpret data quality indicators suggested in the other sections of this document and compare them to production targets.
- Hold feedback sessions with staff involved in data collection and processing
- Conduct “reasonableness” checks by well-informed experts, including pre-release external review in the form of “work in progress”.

14.3.4 Studies of sources of errors

- Studies of sources of errors should be conducted frequently on annual or pluri-annual statistical programs, and conducted occasionally on more frequent programs.
- Consider evaluating, among other sources, errors due to coverage, sampling, non-response, measurement and processing based on the results of studies conducted in other survey steps.

14.4 Overall quality indicators

Quality indicators have been suggested in each of the preceding chapters to measure characteristics specifically related to the topic of the chapter. However it is also suggested that there be some measures related to the project as a whole, not associated with any one particular step. These indicators often cannot be measured until the product has been released and, in some cases, not until much later following the release. Indicators not available until after the release of the product cannot be included in the documentation of the product, but can serve as an indication of the potential quality of a subsequent iteration of the program or of a similar program. These indicators could include the following:

14.4.1 Timeliness

- How long did the project take from start to finish? How long is this from the reference period?
- How long after collection did the estimates for the main characteristics become available?

14.4.2 Relevance

- Do the results respond to the goals of the project and the analytical needs of the community?
- Were there any operational steps or constraints which meant that certain populations may not have been included or certain questions could not be asked?
- Contrast planned outcomes and realized outcomes; justify discrepancies

14.4.3 Interpretability

- Review the completeness of the documentation
- Track the number of requests for information, specifically related to clarifications of the information. This is especially important for repeated surveys. Identify whether this reflects a fundamental flaw in the conceptual framework or in the available documentation

14.4.4 Accuracy

- Was the project able to produce estimates of the desired quality for all of the domains and variables that were planned? This could be expressed as a percentage, e.g. 86% of all planned estimates met the CV targets
- For repeated surveys, compare key estimates and their quality (CV) to previous results. Be able to explain changes. Express changes in the CV in terms of percentage higher/lower than previous iterations. Similar statistics can be generated for imputation rates, error rates etc.
- For non-repeated surveys, potentially use related administrative data or other survey estimates for comparisons to the actual estimates. Since the populations may be somewhat different, explanations of why there are differences may be necessary.

14.4.5 Coherence

- Review reasons for differences in results from previous iterations and try to quantify them (e.g., “The survey now covers the Territories. If they had not been included as in previous iterations, the national estimate would have been 31.4% rather than 31.5%”).
- Review the survey results and those of external sources; address discrepancies

14.4.6 Accessibility

- Provide a description of the types and formats of products from the survey
- Report the number of times a survey product was viewed or accessed on a publicly accessible internet site
- Indicate if the survey data is available in a Public Use Microdata file, if there are any free data products, and if the data is available in the Research Data Centres.

References

Biemer, P., R.M. Groves, N.A. Mathiowetz, L. Lyberg and S. Sudman (eds.) 1991. *Measurement Errors in Surveys*. New York. Wiley. 760 p.

Biemer, P., L. Lyberg. 2003. *Introduction to Survey Quality*. New York. Wiley. 424 p.

Fuller, W. 1987. *Measurement Error Models*. New York. Wiley. 440 p.

Lessler, J.T. and W.D. Kalsbeek. 1992. *Nonsampling Errors in Surveys*. New York. Wiley. 432 p.

Lyberg, L., P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz and D. Trewin (eds.) 1997. *Survey Measurement and Process Quality*. New York. Wiley. 808 p.

Statistics Canada. 2000. "Policy on Informing Users of Data Quality and Methodology." *Statistics Canada Policy Manual*. Section 2.3. Last updated March 4, 2009.
http://icn-rci.statcan.ca/10/10c/10c_010_e.htm.

Statistics Canada. 2002. *Statistics Canada's Quality Assurance Framework - 2002*. Statistics Canada Catalogue no. 12-586-XIE. Ottawa, Ontario. 28 p.
<http://www.statcan.gc.ca/bsolc/olc-cel/olc-cel?lang=eng&catno=12-586-X>.

Statistics Canada. 2003. *Survey Methods and Practices*. Statistics Canada Catalogue no. 12-587-XPE. Ottawa, Ontario. 396 p.
<http://www.statcan.gc.ca/bsolc/olc-cel/olc-cel?lang=eng&catno=12-587-X>.

15 Disclosure control

15.1 Scope and purpose

Disclosure control refers to the measures taken to protect data in accordance with confidentiality requirements. The goal is to ensure that the confidentiality protection provisions are met while preserving the usefulness of the data outputs to the greatest extent possible. Statistics Canada's vigilant disclosure control and confidentiality protection program contributes greatly to data quality. In fact, the high response rates to the Agency's surveys and the public's confidence in it are in a large measure attributable to this.

15.2 Principles

The principles of disclosure control activities are almost entirely governed by the provisions of the Statistics Act (1970, R.S.C. 1985, c. S19), specifically paragraph 17(1)(b):

No person who has been sworn under section 6 shall disclose or knowingly cause to be disclosed, by any means, any information obtained under this Act in such a manner that it is possible from the disclosure to relate the particulars obtained from any individual return to any identifiable individual person, business or organization.

The Statistics Act's confidentiality provisions are extremely rigorous. Consequently, enforcing them in specific cases is a difficult, though extremely important, task. The first goal is to ensure that no identifiable personal information may be inferred within a limited range. Moreover, information must be protected whether or not the subject might be considered confidential by respondents. Finally, how the public perceives the vigilance with which we protect the confidentiality of statistics is at least as important as the actual measures we take to prevent respondents' data from being disclosed.

15.3 Guidelines

15.3.1 General

- Distinguish among the types of data to be processed with each type having its own disclosure control methods. Tabular data are released in the form of statistical tables that are often multi-dimensional. These are further classified as frequency tables and tables of magnitudes. Microdata consist of de-identified records produced for individuals. Finally, some analytical output data can also require disclosure control, particularly if they resemble tabular data (e.g. statistics or histograms) or microdata (e.g. scatter plots or residual values from regressions).
- Check the Disclosure Control Guidelines (long version) to determine which control methods are the most appropriate for your types of data. Limited access methods include access to data from identified data centres, secure remote access and limited access under licence contracts. Limited release methods protect the data itself by reducing or perturbing information.
- Do not disclose the parameters and rules used to control disclosure. Knowing these parameters can help determine more accurately the values of certain respondents.
- Always remember that apparent disclosure can sometimes be just as harmful to the Agency as actual disclosure.

15.3.2 Residual disclosure

- Consider the risk of residual disclosure. This occurs when confidential data can be estimated by cross-referencing released information with other accessible information, including previous releases by the Agency.
- In tables, it is sometimes necessary to find complementary cells to suppress to protect confidential cells. Zero-frequency cells can also pose an attribute disclosure problem, since they eliminate certain possibilities (for example, a zero frequency for the “has a job” category). Often, simply suppressing confidential cells is not sufficient when the marginal totals are also released because it may be possible to calculate the exact value of suppressed cells by solving a system of linear equations. Even if that is not possible, one may derive a range of values for suppressed cells using linear programming methods, and that range might be deemed to be too narrow to sufficiently protect the suppressed value.
- Check whether the categories and hierarchies used by the tables overlap. For example, publishable regions can be subtracted from larger regions, resulting in the publication of a region whose values should be confidential.
- Residual disclosure also occurs when confidential data can be estimated by cross-referencing released information with other accessible information, including previous releases by the Agency. It is hard to define rules to prevent disclosure by cross-reference when several products are released from the same database, particularly in the case of ad hoc requests or output from data centres. Manual intervention is sometimes required. If data can be released from several centres, releases must be coordinated or at least common release rules must be established.

15.3.3 Microdata

- Consider disclosure control methods that are appropriate to microdata dissemination. Data reduction methods include sampling, broadening variable categories (in the case of certain identifiable groups, ensure that the population is large enough), top and bottom coding, removing certain variables from some or all respondents and suppressing some respondents in the file. Data modification methods include adding random noise to the microdata, swapping data, replacing individual values in small groups with average values or deleting information from certain respondents and replacing it with imputed values.
- In longitudinal surveys, define an appropriate strategy before the survey ends. Strategies for releasing microdata files from longitudinal surveys pose an even stickier problem. The strategy must be developed before all the results of the survey are available, i.e. before data is collected for future waves of the survey. Since one of this strategy’s objectives is to define the variables to be released and categorize them, certain assumptions must be made about how those variables evolve over time, particularly if some of them might become key variables.
- In the case of follow-up or second phase surveys, if the main survey has released or plans to release a microdata file, ensure that the microdata file does not pose any additional risks by allowing a composite file to be created by linking the microdata from the two surveys. Assess the success rate achieved by linking the two files and, if it is high, the risk posed by such linkage (for example, what are the consequences of adding identification variables from one survey to the other).
- In accordance with the Policy on Microdata Release (Statistics Canada, 1987) ensure that the Microdata Release Committee reviews all public use microdata files.

15.3.4 Disclosure of certain types of information

- See subsection 17(2) of the Statistics Act, which provides that certain types of confidential information may be released at the discretion of the Chief Statistician and by order. The most common types of such releases are lists of businesses with their addresses and industrial classifications or information related to respondents who have given their consent in writing (waiver). The release of information under the Chief Statistician’s discretionary powers is governed by the Discretionary Release Policy (Statistics Canada, 2004) and, in some cases, by the Guidelines on the Release of Unscreened Microdata under the data sharing agreements described in section 12 or the Act’s discretionary information release provisions.

15.3.5 Resources

- See the confidentiality resources available at Statistics Canada:
 - The Data Access and Control Services Division provides opinions and advice on policies related to the confidentiality of the information collected by Statistics Canada;
 - The Confidentiality and Legislation Committee and its subcommittees, the Disclosure Avoidance Review Group and the Microdata Release Committee provide disclosure control strategies and practices;
 - The Business Survey Methods Division's Disclosure Control Resource Centre provides technical assistance as well as the generalized systems support team for the CONFID software package.
- Use a generalized disclosure control software package such as CONDID rather than customized systems. Such systems reduce the risk of implementation and execution error, the risk of disclosure and the risk of “overprotecting” data, while reducing development costs and times.

15.4 Quality indicators

In general, disclosure control measures reduce data quality by suppressing data or changing detail levels. Disclosure control can also result in access to data being limited to certain groups such as researchers. Certain methods such as data perturbation can affect the accuracy of information released. Bias might arise from value rounding or noise addition.

It is impossible to guarantee absolute confidentiality. Disclosure control is quite complicated and the rules used to measure the extent of the protection provided are somewhat subjective. Although there is no consensus on quality measures, risk functions and loss functions are found primarily.

A loss function measures the extent of the difference between the original data and the data after disclosure control methods have been applied. For altered data (e.g. perturbation), the relative difference between the data before and after adjustment for confidentiality is measured. In the case of suppressed data, the suppression rate indicating the number of values suppressed compared to those released is often used. These indices must be produced at different detail levels and for various respondent groups (e.g. to identify the industrial groups most affected by the suppression).

To a certain extent, a risk function indicates the risk of identifying respondents or values associated with them. In general, for data suppressed in tables, the number of suppressed cells for which protection is inadequate – i.e. a too accurate approximation of the suppressed value can be obtained using information from other cells – must be identified. In the case of microdata, methods tend to measure the risk of disclosure using the re-identification method for a set of characteristic variables (called key variables) or by measuring matching attempts with an external file. Overall, the technique consists of identifying unique combinations of the population found in the released dataset.

References

- Brackstone, G. and P. White. 2002. "Data stewardship at Statistics Canada." *Proceedings of the Social Statistics Section*. American Statistical Association. p. 284-293.
- Doyle, P., J. Lane, J. Theeuwes and L. Zayatz (eds.) 2001. *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*. North-Holland. 462 p.
- Elliot, M., A. Hundepool, E. Schulte Nordholt, J.L. Tambay and T. Wende. 2005. *Glossary on Statistical Disclosure Control*. <http://neon.vb.cbs.nl/casc/glossary.htm>.
- Federal Committee on Statistical Methodology. 2005. *Report on Statistical Disclosure Limitation Methodology*, Statistical Policy Working Paper 22, Second version. Office of Management and Budget. Washington, D.C.
- Hundepool, A., et al. 2008a. *τ -ARGUS version 3.3 User's Manual*. Voorburg. Statistics Netherlands.
- Hundepool, A., et al. 2008b. *μ -ARGUS version 4.2 User's Manual*. Voorburg. Statistics Netherlands.
- Hundepool, A. et al. 2009. *Handbook on Statistical Disclosure Control, Version 1.1*. ESSNet SDC.
- Statistics Canada. 1970. *The Statistics Act*. Ottawa, Canada.
- Statistics Canada. 1987. "Policy on Microdata Release." *Statistics Canada Policy Manual*. Section 4.2. Last updated March 4, 2009. http://icn-rci.statcan.ca/10/10c/10c_026_e.htm.
- Statistics Canada. 2004. "Discretionary Release Policy." *Statistics Canada Policy Manual*. Section 4.3. Last updated March 4, 2009. http://icn-rci.statcan.ca/10/10c/10c_027_e.htm.
- UN Economic Commission for Europe. 2007. *Managing Statistical Confidentiality and Microdata Access – Principles and Guidelines of Good Practice*. United Nations, Geneva.
- Willenborg, L. and T. de Waal. 1996. *Statistical Disclosure Control in Practice*. Springer Verlag. Lecture Notes in Statistics. Vol. 111.
- Willenborg, L. and T. de Waal. 2000. *Elements of Statistical Disclosure Control*. Springer Verlag. Lecture Notes in Statistics. Vol. 155.

16 Data dissemination and communication

16.1 Scope and purpose

Dissemination is the release of data obtained from a statistical activity to users through various media. For each data release there is also a need to effectively communicate the data to data users and a requirement to make known the availability of the release. The Statistics Canada website (www.statcan.ca) is the Agency's principal dissemination channel. The Agency also releases its information in other formats designed to suit the needs of particular users. One release for a single statistical program may include an article in *The Daily*, data tables (e.g. CANSIM or Summary tables), publications, electronic publications and metadata all made available at the same time. Dissemination can also take the form of microdata, a response to a special request, public speech, presentation or television or radio interview.

16.2 Principles

The objectives of dissemination and related communications activities are to maximize the use of Statistics Canada information and to ensure relevance of Statistics Canada by:

- responding to user needs when developing and disseminating information;
- increasing access to information by disseminating directly and through other organizations; and
- providing maximum access to information of broad interest, free of charge, while recovering the costs of providing specialized information and of sustaining an appropriate delivery infrastructure.

These objectives are fundamental to communicating the relevance of Statistics Canada's activities to Canadian households, businesses, institutions, other statistical agencies, other federal government departments, and to the provinces and territories, and to obtain their support for the Agency's collection activities. Most of these objectives are achieved by Statistics Canada's Policy on Dissemination, Communications and Marketing Services (Statistics Canada, 1985). The following guidelines also refer to several other related policies.

16.3 Policies and guidelines

16.3.1 Release of statistical information

- Statistics Canada will authorize the release of microdata files for public use when: (a) the release substantially enhances the analytic value of the data collected; and (b) the Agency is satisfied that all reasonable steps have been taken to prevent the identification of particular survey units (Policy on Microdata Release; Statistics Canada, 1987).
- As outlined in the Policy on Media Relations (Statistics Canada, 2003a), it is the Agency's policy to accept media requests for interviews and to provide comments and data interpretation. Off-the-record background briefings or interviews are not permitted under any circumstance.
- From time to time, erroneous statements about Statistics Canada and its programs or policies, or misinterpretations of data, may appear in the media. In these events, the Agency shall promptly assess the impact of the error and determine the most effective approach by which to respond. Survey managers are encouraged to communicate with the appropriate staff in the Communications and Library Services Division for media training and assistance should it be necessary to clarify a media story.
- Where data validation by an external organization is necessary and where significant benefits to data quality are anticipated or have been previously demonstrated, unreleased non-confidential information may be provided to external organizations for purposes of validation before its official release in *The Daily*, under conditions laid down in the Policy on *The Daily* and Official Release (Statistics Canada, 2008a).
- The production and publication of estimates with future reference dates (frequently referred to as projections or forecasts) is a legitimate part of Statistics Canada's mandate. Such estimates must be released in accordance with the Policy on Estimates with Future Reference Dates (Statistics Canada, 2004a).

16.3.2 Preparation

- All publications that present statistical information and/or analytical findings shall contain a “highlights” section (Policy on Highlights of Publications; Statistics Canada, 2004b).
- Preparation of data to be released from a statistical activity’s source file usually involves many steps. Verify and ensure that released data, after all the processing steps, are consistent with the source data obtained. In the case of regrouped data or derived variables this means that one should be able to reproduce the same results from the source data.
- Thoroughly review all releases (and underlying products) prior to publication from the perspectives of soundness of the data and analysis, appropriateness of the treatment, appropriateness for publication by the Agency, and communications effectiveness.
- Automated tools such as Smart Publishing and automated text comparison software should be used where appropriate to reduce the risk of human error.
- Avoid, as much as possible, preparing the products (i.e. preliminary drafts) while still processing data.
- Develop a dissemination product consistent in style and formatting to other Statistics Canada products: this will assist in its use. Articles written to summarize key findings, trends and contextual information for a general audience should be written according to the Guidelines on Writing for The Daily (available at the Statistics Canada intranet site: http://icn-rci.statcan.ca/02/02_003_e.htm).
- All statistical products must be accompanied by or make explicit reference to documentation on quality and methodology. This documentation must make available to users indicators of the quality of data and descriptions of the underlying concepts and methodology (Policy on Informing Users of Data Quality and Methodology; Statistics Canada, 2000).

16.3.3 Verification

- Have written products reviewed by someone who was not involved in the statistical activity.
- Thoroughly double-check numbers, reference periods (e.g. “in the last six months”, “compared to last quarter”) and words that depict trends (e.g. “increase”, “drop”) in articles and publications to make sure they are accurate.
- Avoid repeating numbers provided in tables in the text; otherwise make sure they are the same.
- Verify numbers in articles and publications against those provided in other tabular products (e.g. CANSIM and Summary tables).
- Test all links in an electronic product before release to ensure that they perform as planned.
- Products are to be disseminated simultaneously in both languages (Official Languages Policy; Statistics Canada, 2004c). Ensure that text in both languages is of high quality and that data and text are consistent in both languages. An automated text comparison tool is available to authors on Statistics Canada’s intranet (Statistics Canada, 2008b).
- All information products, and especially interpretative, analytical and methodological products, for which Statistics Canada is solely or jointly responsible, are subject to review prior to release outside the Agency. The review should ensure that their content is compatible with the Agency’s mandate as a government statistical agency, and that they adhere to the generally accepted norms of good professional practice (Policy on the Review of Information Products (Institutional and Peer Review); Statistics Canada, 2003b).

16.4 Quality indicators

Main quality elements: accessibility, timeliness, relevance

- Describe the availability of products at different levels of detail, formats and media. This indicates whether the statistical information is accessible to a variety of users and for a variety of needs.
- Report the time lag between announcement of the release date and the release of products. Announcing release dates for products in advance enables equal access to all users.
- Report the time lag from the reference date or period to the release of the product. This indicates whether the product is timely with respect to users' needs.
- Report the time lag between scheduled release date and actual release date. This is a measure of the punctuality of the product.
- Document the occurrence of errors detected during the editing step just prior to the release of products. Errors detected so late in the production of statistical information indicate a higher risk of having to correct other errors after release.
- Document the occurrence of errors detected after release.
- Monitor how frequently information products are accessed by users over time. A decline in the product's relevance could be indicated by a decline in user activity.

References

Statistics Canada. 1987. "Policy on Microdata Release." *Statistics Canada Policy Manual*. Section 4.2. Last updated March 4, 2009.

http://icn-rci.statcan.ca/10/10c/10c_026_e.htm.

Statistics Canada. 2000. "Policy on Informing Users of Data Quality and Methodology." *Statistics Canada Policy Manual*. Section 2.3. Last updated March 4, 2009.

http://icn-rci.statcan.ca/10/10c/10c_010_e.htm.

Statistics Canada. 2003a. "Policy on Media Relations: Spokespersons and Response to the Media." *Statistics Canada Policy Manual*. Section 1.2. Last updated March 4, 2009.

http://icn-rci.statcan.ca/10/10c/10c_002_e.htm.

Statistics Canada. 2003b. "Policy on the Review of Information Products (Institutional and Peer Review)." *Statistics Canada Policy Manual*. Section 2.5. Last updated March 4, 2009.

http://icn-rci.statcan.ca/10/10c/10c_011_e.htm.

Statistics Canada. 2004. "Policy on Dissemination, Communications and Marketing Services." *Statistics Canada Policy Manual*. Section 3.1. Last updated March 4, 2009.

http://icn-rci.statcan.ca/10/10c/10c_015_e.htm.

Statistics Canada. 2004a. "Policy on Estimates with Future Reference Dates." *Statistics Canada Policy Manual*. Section 2.2. Last updated March 4, 2009.

http://icn-rci.statcan.ca/10/10c/10c_009_e.htm.

Statistics Canada. 2004b. "Policy on Highlights of Publications." *Statistics Canada Policy Manual*. Section 2.1. Last updated March 4, 2009.

http://icn-rci.statcan.ca/10/10c/10c_008_e.htm.

Statistics Canada. 2004c. "Official Languages Policy." *Statistics Canada Policy Manual*. Section 5.3. Last updated March 5, 2009.

http://icn-rci.statcan.ca/10/10c/10c_034_e.htm.

Statistics Canada. 2008. "Policy on The Daily and Official Release." *Statistics Canada Policy Manual*. Section 3.2. Last updated March 5, 2009.

http://icn-rci.statcan.ca/10/10c/10c_058-eng.htm.

17 Data analysis and presentation

17.1 Scope and purpose

Data analysis is the process of developing answers to questions through the examination and interpretation of data. The basic steps in the analytic process consist of identifying issues, determining the availability of suitable data, deciding on which methods are appropriate for answering the questions of interest, applying the methods and evaluating, summarizing and communicating the results.

Analytical results underscore the usefulness of data sources by shedding light on relevant issues. Some Statistics Canada programs depend on analytical output as a major data product because, for confidentiality reasons, it is not possible to release the microdata to the public. Data analysis also plays a key role in data quality assessment by pointing to data quality problems in a given survey. Analysis can thus influence future improvements to the survey process.

Data analysis is essential for understanding results from surveys, administrative sources and pilot studies; for providing information on data gaps; for designing and redesigning surveys; for planning new statistical activities; and for formulating quality objectives.

Results of data analysis are often published or summarized in official Statistics Canada releases.

17.2 Principles

A statistical agency is concerned with the relevance and usefulness to users of the information contained in its data. Analysis is the principal tool for obtaining information from the data.

Data from a survey can be used for descriptive or analytic studies. Descriptive studies are directed at the estimation of summary measures of a target population, for example, the average profits of owner-operated businesses in 2005 or the proportion of 2007 high-school graduates who went on to higher education in the next twelve months. Analytic studies may be used to explain the behaviour of and relationships among characteristics; for example, a study of risk factors for obesity in children would be analytic.

To be effective, the analyst needs to understand the relevant issues (both current and those likely to emerge in the future) and how to present the results to the audience. Study of background information allows the analyst to choose suitable data sources and appropriate statistical methods. Any conclusions presented in an analysis, including those that can impact public policy, must be supported by the data being analyzed.

17.3 Guidelines

17.3.1 Initial preparation

Prior to conducting an analytic study the following questions should be addressed:

- Objectives. What are the objectives of this analysis? What issue am I addressing? What question(s) will I answer?
- Justification. Why is this issue interesting? How will these answers contribute to existing knowledge? How is this study relevant?
- Data. What data am I using? Why is it the best source for this analysis? Are there any limitations?
- Analytical methods. What statistical techniques are appropriate? Will they satisfy the objectives?
- Audience. Who is interested in this issue and why?

17.3.2 Suitable data

- Ensure that the data are appropriate for the analysis to be carried out. This requires investigation of a wide range of details such as whether the target population of the data source is sufficiently related to the target population of the analysis, whether the source variables and their concepts and definitions are relevant to the study, whether the longitudinal or cross-sectional nature of the data source is appropriate for the analysis, whether the sample size in the study domain is sufficient to obtain meaningful results and whether the quality of the data, as outlined in the survey documentation or assessed through analysis is sufficient.
- If more than one data source is being used for the analysis, investigate whether the sources are consistent and how they may be appropriately integrated into the analysis.

17.3.3 Appropriate methods and tools

- Choose an analytical approach that is appropriate for the question being investigated and the data to be analyzed.
- When analyzing data from a probability sample, analytical methods that ignore the survey design can be appropriate, provided that sufficient model conditions for analysis are met. (See Binder and Roberts, 2003.) However, methods that incorporate the sample design information will generally be effective even when some aspects of the model are incorrectly specified.
- Assess whether the survey design information can be incorporated into the analysis and if so how this should be done – such as by using design-based methods. See Binder and Roberts (2009) and Thompson (1997) for discussion of approaches to inferences on data from a probability sample.
 - See Chambers and Skinner (2003), Korn and Graubard (1999), Lehtonen and Pahkinen (1995), Lohr (1999), and Skinner, Holt and Smith (1989) for a number of examples illustrating design-based analytical methods.
 - For a design-based analysis consult the survey documentation about the recommended approach for variance estimation for the survey. If the data from more than one survey are included in the same analysis, determine whether or not the different samples were independently selected and how this would impact the appropriate approach to variance estimation.
 - The data files for probability surveys frequently contain more than one weight variable, particularly if the survey is longitudinal or if it has both cross-sectional and longitudinal purposes. Consult the survey documentation and survey experts if it is not obvious as to which might be the best weight to be used in any particular design-based analysis.
 - When analyzing data from a probability survey, there may be insufficient design information available to carry out analyses using a full design-based approach. Assess the alternatives.
- Consult with experts on the subject matter, on the data source and on the statistical methods if any of these is unfamiliar to you.
- Having determined the appropriate analytical method for the data, investigate the software choices that are available to apply the method. If analyzing data from a probability sample by design-based methods, use software specifically for survey data since standard analytical software packages that can produce weighted point estimates do not correctly calculate variances for survey-weighted estimates.
- It is advisable to use commercial software, if suitable, for implementing the chosen analyses, since these software packages have usually undergone more testing than non-commercial software.
- Determine whether it is necessary to reformat your data in order to use the selected software.
- Include a variety of diagnostics among your analytical methods if you are fitting any models to your data.

- Data sources vary widely with respect to missing data. At one extreme are the sources which seem complete - where any missing units have been accounted for through a weight variable with a non-response component and all missing items on responding units have been filled in by imputed values. At the other extreme are data sources where no processing has been done with respect to missing data. The work required by the analyst to handle missing data can thus vary widely. It should be noted that the handling of missing data in analysis is an ongoing topic of research.
 - Refer to the documentation about the data source to determine the degree and types of missing data and the processing of missing data that has been performed. This information will be a starting point for what further work may be required.
 - Consider how unit and/or item non-response could be handled in the analysis, taking into consideration the degree and types of missing data in the data sources being used.
 - Consider whether imputed values should be included in the analysis and if so, how they should be handled. If imputed values are not used, consideration must be given to what other methods may be used to properly account for the effect of non-response in the analysis.
 - If the analysis includes modelling, it could be appropriate to include some aspects of non-response in the analytical model.
 - Report any caveats about how the approaches used to handle missing data could have impact on results.

17.3.4 Interpretation of results

- Since most analyses are based on observational studies rather than on the results of a controlled experiment, avoid drawing conclusions concerning causality.
- When studying changes over time, beware of focusing on short-term trends without inspecting them in light of medium- and long-term trends. Frequently, short-term trends are merely minor fluctuations around a more important medium- and/or long-term trend.
- Where possible, avoid arbitrary time reference points. Instead, use meaningful points of reference, such as the last major turning point for economic data, generation-to-generation differences for demographic statistics, and legislative changes for social statistics.

17.3.5 Presentation of results

- Focus the article on the important variables and topics. Trying to be too comprehensive will often interfere with a strong story line.
- Arrange ideas in a logical order and in order of relevance or importance. Use headings, sub-headings and sidebars to strengthen the organization of the article.
- Keep the language as simple as the subject permits. Depending on the targeted audience for the article, some loss of precision may sometimes be an acceptable tradeoff for more readable text.
- Use graphs in addition to text and tables to communicate the message. Use headings that capture the meaning (e.g., “Women’s earnings still trail men’s”) in preference to traditional chart titles (e.g., “Income by age and sex”). Always help readers understand the information in the tables and charts by discussing it in the text.
- When tables are used, take care that the overall format contributes to the clarity of the data in the tables and prevents misinterpretation. This includes spacing; the wording, placement and appearance of titles; row and column headings and other labeling.
- Explain rounding practices or procedures. In the presentation of rounded data, do not use more significant digits than are consistent with the accuracy of the data.

- Satisfy any confidentiality requirements (e.g. minimum cell sizes) imposed by the surveys or administrative sources whose data are being analysed.
- Include information about the data sources used and any shortcomings in the data that may have affected the analysis. Either have a section in the paper about the data or a reference to where the reader can get the details.
- Include information about the analytical methods and tools used. Either have a section on methods or a reference to where the reader can get the details.
- Include information regarding the quality of the results. Standard errors, confidence intervals and/or coefficients of variation provide the reader important information about data quality. The choice of indicator may vary depending on where the article is published.
- Ensure that all references are accurate, consistent, and are referred to in the text.
- Check for errors in the article. Check details such as the consistency of figures used in the text, tables and charts, the accuracy of external data, and simple arithmetic.
- Ensure that the intentions stated in the introduction are fulfilled by the rest of the article. Make sure that the conclusions are consistent with the evidence.
- Have the article reviewed by others for relevance, accuracy and comprehensibility, regardless of where it is to be disseminated. As a good practice, ask someone from the data providing division to review how the data were used. If the article is to be disseminated outside of Statistics Canada, it must undergo institutional and peer review as specified in the Policy on the Review of Information Products (Statistics Canada, 2003).
- If the article is to be disseminated in a Statistics Canada publication make sure that it complies with the current Statistics Canada Publishing Standards. These standards affect graphs, tables and style, among other things.
- As a good practice, consider presenting the results to peers prior to finalizing the text. This is another kind of peer review that can help improve the article. Always do a dry run of presentations involving external audiences.
- Refer to available documents that could provide further guidance for improvement of your article, such as Guidelines on Writing Analytical Articles (Statistics Canada 2008) and the Style Guide (Statistics Canada 2004)

17.4 Quality indicators

Main quality elements: relevance, interpretability, accuracy, accessibility

An analytical product is relevant if there is an audience who is (or will be) interested in the results of the study.

For the interpretability of an analytical article to be high, the style of writing must suit the intended audience. As well, sufficient details must be provided that another person, if allowed access to the data, could replicate the results.

For an analytical product to be accurate, appropriate methods and tools need to be used to produce the results.

For an analytical product to be accessible, it must be available to people for whom the research results would be useful.

References

- Binder, D.A. and G.R. Roberts. 2003. "Design based methods for estimating model parameters." In *Analysis of Survey Data*. R.L. Chambers and C.J. Skinner (eds.) Chichester. Wiley. p. 29-48.
- Binder, D.A. and G. Roberts. 2009. "Design and Model Based Inference for Model Parameters." In *Handbook of Statistics 29B: Sample Surveys: Inference and Analysis*. Pfeiffermann, D. and Rao, C.R. (eds.) Vol. 29B. Chapter 24. Amsterdam. Elsevier. 666 p.
- Chambers, R.L. and C.J. Skinner (eds.) 2003. *Analysis of Survey Data*. Chichester. Wiley. 398 p.
- Korn, E.L. and B.I. Graubard. 1999. *Analysis of Health Surveys*. New York. Wiley. 408 p.
- Lehtonen, R. and E.J. Pahkinen. 2004. *Practical Methods for Design and Analysis of Complex Surveys*. Second edition. Chichester. Wiley.
- Lohr, S.L. 1999. *Sampling: Design and Analysis*. Duxbury Press. 512 p.
- Skinner, C.K., D.Holt and T.M.F. Smith. 1989. *Analysis of Complex Surveys*. Chichester. Wiley. 328 p.
- Thompson, M.E. 1997. *Theory of Sample Surveys*. London. Chapman and Hall. 312 p.
- Statistics Canada. 2003. "Policy on the Review of Information Products." *Statistics Canada Policy Manual*. Section 2.5. Last updated March 4, 2009. http://icn-rci.statcan.ca/10/10c/10c_011_e.htm.
- Statistics Canada. 2004. *Style Guide*. Last updated October 6, 2004. http://icn-rci.statcan.ca/10/10d/10d_000_e.htm.
- Statistics Canada. 2008. *Guidelines on Writing Analytical Articles*. Last updated September 16, 2008. http://icn-rci.statcan.ca/10/10g/10g_001_e.htm.

18 Documentation

18.1 Scope and purpose

Documentation constitutes a record of the statistical activity, including the concepts, definitions and methods used to collect, process and analyze data and produce statistical products. It is intended to promote effective and informed use of data. Quality indicators produced during a statistical activity should be included in the documentation. Analysis of quality indicators, in terms of their impact on the use of statistical products resulting from statistical activity, should also be included in the documentation.

During implementation, documentation is a means of communication to ensure effective development of a statistical activity. It includes not only what decisions were made, but also why they were made, and provides information that will be useful for future development and implementation of the same statistical activity or a similar or redesigned activity.

18.2 Principles

The goal of documentation is to provide a complete, unambiguous and multi-purpose record of the statistical activity, including its outputs. Documentation may be intended for various target audiences, such as management, technical staff, planners of other surveys, and users. It should be readily accessible, up to date, timely as to ensure relevance, and comprehensible to its main audience. It can be multimedia format (e.g. hardcopy, electronic format, visual presentation). Care must be taken to preserve statistical activity documents.

18.3 Guidelines

18.3.1 Tailor the documentation to the target audience and the general context

- The level of documentation should consider the target audience for which it is intended. As a result, it must be determined whether documentation should be detailed or condensed, technical or general. In cases where statistical products are disseminated by Statistics Canada, the documentation must meet the requirements of the Policy on Informing Users of Data Quality and Methodology (Statistics Canada, 2000).
- The scope of the documentation should take into consideration the context in which the statistical activity concerned was carried out, the importance of the statistical activity, whether it is new or recurrent, and whether it is similar to or different from other statistical activities conducted by the organization. This determines what must be covered by new documentation and what can be covered by references to existing documentation.
- Documentation priorities should also take into account the statistical activity's budget, the appropriate time for publishing the documentation, as well as its short-term and long-term benefits. There must be no delay between completion of the statistical activity and preparation of the documentation. Any such delay may not only affect the documentation's timeliness and relevance, but also erode its accuracy.

18.3.2 Provide complete, accurate documentation

- Statistical activity documentation generally comes in three types: (1) general documents that, for the most part, provide a current picture of the statistical activity, (2) thematic documents that provide details on implementation, and (3) thematic evaluations.

General documentation

- Objectives: Include information on the objectives and uses of the data, timeliness, frequency of the statistical activity, and data quality targets. Objectives can change as the survey progresses (e.g. budgetary constraints, perceived feasibility, results of new pilot studies, or new technology). Such changes must be documented as they have an impact on questionnaire design and on test result analysis.

- **Content:** Include the concepts, definitions and the questionnaire used. To facilitate integration with other sources, use standardized concepts, questions, methods, processes and classifications. Highlight differences, if warranted. Mention the role of advisory committees and users.
- **Methodology:** Deal with issues such as target population, sampling frame, coverage, reference period, sample design, sample size and selection method, collection method and follow-up procedures for non-response, edit and imputation, estimation, benchmarking and revision, seasonal adjustment and confidentiality. Provide a methodological overview. Emphasize various aspects for different readers. Provide a consolidated document of technical issues for technical staff.
- **Data quality:** Provide general information about coverage, sampling error, non-sampling error, response rates, the rates and effects of edit and imputation, comparability over time and with other data, validation studies, quality assurance measures and any other relevant measures specific to the statistical activity concerned. Describe any unexpected events affecting data quality (e.g. flooding, high non-response rate). For technical users, include total variance or its components by source, non-response and response biases, and the impact and interpretation of seasonal adjustment.

Detailed documentation on implementation

- **Activity planning and budget**
- **Operations:** Include an interviewer manual, training manuals, instructions or a manual for supervisors and quality control staff, manuals for data capture and processing staff, and feedback and debriefing reports.
- For computer-assisted interviewing, provide the computer application's development specifications.
- **Systems:** Include information on the data files (record layouts, explanation of codes, basic frequencies, edit procedures), systems documentation (construction, algorithms, use, storage and retrieval) and monitoring reports (time spent on specific activities, trouble areas, scheduling of runs to determine whether processing is on time).
- **Implementation:** Document all operations, with inputs and outputs clearly specified. Attach work schedules for each implementation step.
- **Resources:** List the resources used and when. Provide an account of all salary and non-salary expenditures (amounts and time). Comment on expenditures in relation to budgets.

Evaluations

- Prepare a general evaluation of the statistical activity process.
- Describe cognitive tests, field tests or pilot surveys and report on results and recommendations in relation to specifications.
- Document methodology evaluations, such as alternative sample designs considered or the performance of the sample design used.
- If the documentation is to be disseminated outside Statistics Canada, it must undergo institutional and peer review as specified in the Policy on the Review of Information Products (Statistics Canada, 2003). Even if it is exclusively for internal use, documentation should be reviewed by managers, target audience representatives or peers for relevance, accuracy and comprehensibility.
- Make the documentation accessible
- In the case of documentation intended for users, provide the documentation elements required for the Integrated Metadatabase (Statistics Canada, 2000c). As the archive for information about Statistics Canada's surveys and programs, the IMDB contains most of the information for users regarding methodology and data accuracy. Electronic products contain a link to the IMDB, which is used to access documentation about the product. For print products, the IMDB provides adequate documentation, in accordance with the Policy on Informing Users of Data Quality and Methodology (Statistics Canada, 2000d).

- Choose tools that, as much as possible, provide a central repository for documentation about statistical activities and encourage structured storage and file search. At a minimum, each document should present a clear title, a date and the names of the authors (institutions or individuals). The preservation of Statistics Canada documents is required under the Policy on Document Management.
- Sort and document references (theoretical and general articles and documents concerning the project that were, however, not produced as part of the project).

18.4 Quality indicators

Main quality elements: interpretability, accessibility, timeliness

- Number of documents published externally and that underwent institutional and peer review.
- Published statistical products that satisfy the Policy on Informing Users of Data Quality and Methodology.

References

Statistics Canada. 2000. "Policy on document management." *Statistics Canada Policy Manual*. Section 5.9. Last updated March 5, 2009.
http://icn-rci.statcan.ca/10/10c/10c_040_e.htm.

Statistics Canada. 2000d. "Policy on informing users of data quality and methodology." *Statistics Canada Policy Manual*. Section 2.3. Last updated March 4, 2009.
http://icn-rci.statcan.ca/10/10c/10c_010_e.htm.

Statistics Canada. 2002c. Statistics Canada's Quality Assurance Framework – 2002. Catalogue No. 12-586-XIE.

Statistics Canada. 2003. "Policy on the review of information products." *Statistics Canada Policy Manual*. Section 2.5. Last updated March 4, 2009.
http://icn-rci.statcan.ca/10/10c/10c_011_e.htm.

Statistics Canada. 2004. "Policy on standards." *Statistics Canada Policy Manual*. Section 2.10. Last updated March 4, 2009.
http://icn-rci.statcan.ca/10/10c/10c_014_e.htm

Statistics Canada. 2007. "Integrated Metadatabase – Guidelines for Authors." *Standards Division Internal Communications Network*. No date.
<http://stdsweb/standards/imdb/imdb-menu.htm>

United Nations, Conference of European Statisticians. 1983. *Draft guidelines for the preparation of presentations of the scope and quality of statistics for users*. Geneva, Switzerland.