Journal of Official Statistics, Vol. 17, No. 1, 2001, pp. 1–20

Can a Statistician Deliver?

Richard Platek[1] *and Carl-Erik Särndal*[2]

The theme in this article is quality in the statistics produced by National Statistical Agencies. Good survey methodology is important for realizing quality. Survey methodology has been described as a collection of practices backed by some theory. It is perhaps too optimistic to expect a complete and coherent theory for as vast a field of activities as national statistics production. Yet the desire for a more satisfactory theoretical basis is voiced by many. We discuss progress and failures in survey theory advancement. We note that the different types of statisticians participating in the production process may view quality in possibly different ways; their views, on the other hand, may not wholly agree with those held by the users of statistics.

Key words: Statistics production; official statistics; quality; accuracy; relevance; survey design; survey operations; total survey error.

1. Introduction

National Statistical Agencies have always tried to achieve high quality in the statistics that they produce. This effort has been brought further into focus by the recent emphasis on systematic quality management. One approach is known as Total Quality Management, or TQM.

Several National Statistical Agencies have formulated their own definitions of the nebulous concept of ''quality in (official) statistics,'' for example Statistics Sweden (1994) and Statistics Canada (1998). The very similar quality concept for official statistics given in the Encyclopedia of Statistical Sciences by Elvers and Rosén (1998) has five principal dimensions: (1) Contents, (2) Accuracy, (3) Timeliness, (4) Coherence, especially Comparability, and (5) Availability and Clarity. Each has a number of sub-dimensions.

Any attempt at defining ''quality'' is bound to be challenged; it is utopian to expect universal agreement on what quality is. The definitions of Statistics Sweden and Statistics Canada emphasize the user's needs and concerns in regard to data. User orientation is important in TQM.

The role of survey methodology is to promote high quality in the statistics produced. According to Statistics Canada (1998), survey methodology is ''a collection of practices, backed by some theory and empirical evaluation, among which practitioners have to make sensible choices in the context of a particular application.'' Many may agree with this description; some will feel disappointed. Despite a long history, where at least the last 100 years have revealed a clear scientific ambition, there still seems to be a weak

[1] Consultant, 1125 Ambleside, Ottawa, Ontario, K2B 8E2, Canada.
[2] Consultant, 2115 Erinbrook Crescent no. 44, Ottawa, Ontario, K1B 4J5, Canada. E-mail: carl.sarndal@home.com

theoretical basis for the production of official statistics. This is disappointing, in particular in view of the importance that those statistics can have in high-level decision making, by governments and others.

There is a difference between practice built on firm theory and a collection of practices. The latter has no unifying conceptualization, no complete basis for systematic generalizations. Some of the practices may appeal to theory, but it is piecemeal theory, rather than a holistic view.

Consequently, one and the same operation in the statistical production cycle may be carried out quite differently, perhaps haphazardly, in different statistical agencies around the world. One cannot challenge any particular way of operating, because no firm theory dictates the choice. Ad hoc solutions may be the result. Often, smaller agencies will monitor and adopt the practice of well respected agencies, but ''following the leader'' is not the same as a robust ''guidance by theory.'' Moreover, this puts a heavy burden of responsibility on the leading organizations.

A theme in this article is the contrast between the quest for scientific rigour and cohesiveness on the one hand, and practical demands on the other. We recognize that some other professional fields, such as law and accountancy, are more or less in the same position as official statistics: there is a set of practices, but no complete theory.

Some agencies, such as Westat (see Morganstein and Marker 1997) and Statistics Sweden (see Lyberg, Biemer, and Japec 1998), advocate the use of Current Best Methods (CBM) manuals. In other agencies, a similar document may be called Quality Guidelines, as at Statistics Canada (1998). Such manuals advocate practices which are commensurate with experience and practice in the agency. They can assist statisticians in the design and operation of surveys. Especially if written in a nontechnical language, they may become widely used in the agency. They encourage standardization for ''recurring processes,'' that is, processes common to many surveys (and for which there is some consensus about ''bestness''). For example, a CBM on nonresponse prevention will describe methods thought to be effective, or ''currently best,'' for realizing high response rate and low response error.

The name Current Best Methods is suggestive: optimal procedures for a given process may not be derivable from any commonly accepted theory, but some procedures are considered by experienced professionals to be adequate in today's practice, fully recognizing that they may not be so in tomorrow's practice.

Systematic quality management poses new challenges for the statisticians. Achieving high accuracy, or reducing total error, is an important part of the striving for quality. Total survey error models have been the statistician's attempt to measure the most important components of the total error. The slow progress in total survey error modeling is discouraging to some. Smith (1994, p. 10) notes that ''a systematic plan to reduce the total error is necessary if we are to have total quality management in surveys.'' A few years earlier, the same author, Smith (1990) noted: ''The failure of both theoretical and practical statisticians to integrate sampling and nonsampling errors into measures of total survey error even after 50 years of intensive research must be noted as one of the failures of this important branch of statistics.''

We have written this article following a series of informal discussions about challenges that we see for statisticians in national statistical agencies. The title, Can a statistician deliver?, poses a question that may seem provocative to some. Do we suggest that the statistician cannot properly do his or her job? Who is the statistician alluded to in the title?

''Deliver'' implies that there is a provider as well as a receiving party. The receiver is a user of statistical information, called ''the user'' from now on. Our statistician can belong to any one of the professional categories found in the statistical agency. We return to this identification in the next section.

The user asks: Can I trust the data and the statistics delivered to me? To what extent do they serve my purpose? The user seeks quality assurance. Is the statistician capable of delivering not only numbers but also adequate quality assurance? What forms should the assurance take?

We believe that many feel, with us, that statistical agencies (and the statisticians who work there) do not provide sufficiently explicit answers to these questions. Our primary objective is to point to some aspects of surveys – we use the term here in a wide sense to include sample surveys, censuses, and statistics derived from administrative registers – that may have not received enough attention. The statisticians play an important role, because of the importance of the data they produce, and their task is difficult. We suggest that they do not always ''deliver'' fully; perhaps they should not be expected to do so.

The article is arranged as follows: We note that different types of statisticians participate in the production process (Section 2). We propose that different participants view quality in different lights: the statisticians on the one hand, the users of statistics on the other (Sections 3, 4). Survey errors and quality indicators are discussed (Sections 5, 6). Some recent advancements in survey theory are scrutinized (Section 7). Total survey error and the modeling of this error are discussed (Section 8). Challenges in the further progress of survey theory are noted (Section 9). Some conclusions are drawn (Section 10).

2. Who Is the Statistician?

The statistician in our title is anyone who contributes to the ultimate delivery of statistics and data to users. He or she can belong to any one of the professional categories typically found in a large statistical agency: theoretical statistician, survey methodologist, subject matter specialist, information technologist, and survey manager. All contribute to the production process.

Given our own background, we speak relatively more about methodologists and theoreticians, those whose university training is typically in statistical science, and whose task is to put good statistical machinery to work in order to produce the statistics to be delivered. The survey methodologist constructs (sometimes highly complex) plans for sample selection and estimation; he or she has a strong interest in the accuracy of the statistics produced, and likes to measure it by the design-based variance estimate or coefficient of variation (cv). But questions of survey cost are also important to these statisticians. Some methodologists, in particular those with a psychometrics or sociometry background, focus on important matters relating to respondent cooperation and questionnaires.

The subject matter statisticians, whose training is often in Economics or Sociology, combine extensive knowledge of the survey background with a sound statistical intuition.

One of their important concerns is to examine the statistics before delivery, to ensure that they ''make good sense,'' so that there is coherence with the results published in the recent past. They analyze and draw conclusions from the data. They are likely to be less preoccupied with formal accuracy measurement than the methodologists. The information technology experts play a vital role with regard to the systems aspects of a survey and with regard to making data available in convenient forms. The project managers, another important category, focus on operations management, and often take an active part in survey planning. Ideally at least, they are in frequent contact with users and their needs, so as to ensure a delivery of appropriate data. Timeliness, field control and cost are issues constantly on the mind of the survey manager.

Each category of statisticians has particular commitments and responsibilities within the statistical programs to which they contribute. This is natural. To illustrate, let us look at the theoretician, someone trained in statistical science, who may be an academic working as a consultant to the statistical agency, or employed full time by the agency. He or she is expected to create and/or apply new methodology and to suggest modification to existing methodology. He or she tends to take a general approach to a problem, and the solutions sometimes lie outside the immediate needs of the survey. Vital aspects of data quality such as relevance, comparability, timeliness, and forms of dissemination fall mainly outside the traditional concerns of the theoretician, who may see little mathematical and technical challenge in those aspects of quality. Thus the theoretician's contribution to data quality is, despite the high quality of this technical input, somewhat incomplete. He or she may not be fully aware of the narrow focus of the contribution, in part because a myopic view of a survey is often reinforced by scientific and professional journals, whose publication policy may encourage concise theoretical work. But still, the theoretician's effort is important; regrettably it is often less visible and less valued by the organization than that of some participants closer to operations.

What we have just said about a narrow focus among the theoreticians holds for other categories of statisticians as well. They focus on the respective issues that concern them the most. This is not a negative aspect; the different participants should bring their own personal competence to bear on the joint production effort.

Communication is essential between the different statisticians working together on a survey. For this reason, statistical agencies rely on project teams with representation from all or most of these categories. This is one way of assuring that the views and the competence of all essential groups are taken into account. The methodologist's task is to see to it that theory is put to good use. The subject matter statistician's expertise on the survey background gives him or her a solid basis for interpreting the results. The information technology expert focuses on the proper functioning of systems and data flows. The project manager makes sure that the whole survey process runs smoothly. Collectively, these statisticians must satisfy the user. The success of a survey is determined by how well the different contributions are reconciled and coordinated and by the extent to which the user will sense that a competent professional team lies behind the published statistics. The statisticians have a collective responsibility to justify and explain the methods, the survey processes and the results to the user.

3. What Is Data Quality?

What is data quality? A number of answers have been proposed. Surveys can always be made better, and the perceived importance, the frequency and the understanding of different imperfections will colour an individual statistician's attitude: quality will mean different things to different statisticians, irrespective of what the agency's official quality concept states. Each category views quality from its own vantage point, molded by background, university education and particular task in the production process.

(i) To the theoretical statistician, an important element of quality is to be able to put the most recent advances in statistical science to work, in proposing some new technique for sample selection or for estimation, the goal being to improve the accuracy of the estimates.
(ii) To the survey methodologist, quality is reflected in the overall survey design, including the size of variances and cv's of the survey estimates. He or she wants to maintain high response rates and apply good imputation procedures, while respecting the cost frame for the survey.
(iii) To the subject matter specialist, the meaning and the presentation of data are important aspects of quality. He or she seeks clarity in the definition of concepts and coherence in the presentation of the data, and strives to provide the users with competent advice on the applicability and the limitations of the data.
(iv) To the production manager, quality is realized if the production succeeds in respecting given cost frames and aspects of timeliness.
(v) To the informatics specialist, a dominating concern is the smooth functioning of the computerized systems for data collection, edit, imputation, and compilation, and he or she wants to make sure that data bases and other outputs can be easily accessed.

Despite a physical proximity in the same statistical agency, these statisticians live to some extent in separate worlds. For example, a survey manager may be aware that the procedures for nonresponse treatment need to be revised and updated, in ways advocated by the theoreticians and the methodologists, but considerations of time and cost may for the time being rule out an implementation of new procedures.

Each statistician contributing to the survey is, despite team meetings, centered on his or her immediate responsibilities. It is hard to make them feel equally strongly about all of the abstract components of the agency's official quality concept. It is imperative that the statisticians are not semi-skilled, but highly trained and motivated individuals. The average citizen's life is greatly affected by statistical data; data on economic, social, medical, and legal issues have become fundamental to everybody's existence and well-being. Statisticians must be aware both of the high responsibility associated with their respective contributions, and of the fact that society expects a lot from them.

The user's perspective is quite different: It is a view from the outside. Released survey results are of interest if they satisfy elements of the user's own personal idea of quality and usefulness. He or she will ask questions such as: (a) are the results comparable with, or consistent with, numerical evidence from different but related data sources? (b) is the information timely? (c) does it compare well with previous observations, does it fluctuate credibly over time? (d) can I trust the measurement procedure, the nonresponse adjustments and the estimation procedure behind the results? (e) are the data made available in a manner that is easily understood, meaningful and useful for my purposes? If these questions mirror the user's view of quality, it is clear that they represent a view different from at least some of the statisticians. The user's view may be fairly close to that of the subject matter statistician but quite far from that of the theoretician, the methodologist and the informatics specialist. Moreover, the user cannot, or prefers not to, question accuracy; the view from the outside simply does not invite a close scrutiny of accuracy.

The user sometimes expects too much from, or reads too much into, the data delivered. A trained statistician views a published statistic as an estimate of some characteristic of a finite population, at a given point in time, namely the time of the survey. On the other hand, the user may, perhaps subconsciously, view the information as an extrapolating statement about a future date. Abstract concepts, such as that of a finite population of identifiable units existing at a fixed reference point in time, may be elusive to many users, and the statisticians may be at fault for not making such basic concepts abundantly clear. Although they strive to be clear, the goal may not always be realized.

Increased focus on perceived user needs has in recent years inspired several national statistical agencies to elaborate their own definitions of ''quality.'' Fitness for use, and user orientation are dominating themes in these quality definitions. It is a positive result that the user now receives more attention than in the past, when quality was interpreted as being more or less synonymous with accuracy, which some would call ''the statistical quality'' because of its roots in random variation.

The trend toward a more tangible user satisfaction is not without side effects. It shifts the focus (inside the agency) to satisfying those aspects of quality that users (outside the agency) can more readily observe and identify with, such as timeliness, comparability over time, relevance, availability, and clarity. By contrast, accuracy is not one of these.

We argued that individual statisticians focus on those components of quality which agree most closely with their personal experience of good workmanship. Accuracy is central to the methodologists and the theoreticians. It is crucial that the organization maintains a staff of specialists on accuracy, where we should understand accuracy in the broad sense, not only in the sense of sampling variance. Accuracy is determined in large part by the survey design, that is, the preconceived plan about sampling, questionnaires, data collection, imputation, and estimation.

Accuracy is important to the user, too. Without an assurance of reasonably accurate data, the user will not have confidence. A realistic user recognizes that a relatively large sampling variance in some of the estimates does not condemn all aspects of a survey, just as small sampling variance does not by itself ensure ''good data.'' He or she knows that the accuracy depends on the targeted population level, making estimates for small sub-populations (small areas) have a large sampling variance. The discriminating user does not rely only on sampling variance for his or her assessment of quality; he or she will ask questions about other vital aspects of the survey, namely the survey design and the survey process. He or she is aware that poor quality may have its roots in underestimated total cost of the survey, insufficient control of the survey operations, or improper allocation of cost to operations.

In recent times, survey processes have been revolutionized by the progress in information technology. We have experienced more efficient procedures for data collection, for finalizing the data set, for estimation, tabulation and dissemination. Shorter production times have been realized. But it is not obvious that one can convince users that improvements on all those counts, important as they may be, are accompanied by significant improvement in overall data quality. In particular, accuracy may not be improved. How do we as statisticians define, control and measure the many operations entering into the whole survey process, so that the user can feel properly informed? The task is not simple, yet an important part of the statistician's responsibility lies in providing information of this kind.

4. Design Versus Process

A useful distinction is survey design versus survey process. It is different from the distinction ''research culture'' versus ''operations culture'' used in Dillman's (1996) analysis of innovation in statistical agencies. Discussants of Dillman's article propose other dichotomies that they see as perhaps more relevant; Colledge's (1996) distinction, ''survey design including applied research'' versus ''survey operations,'' seems to come close to ours. All such distinctions are to some degree arbitrary; nevertheless, they can enlighten a discussion. Design is conceptualization, process is the execution of the design. Design and process often involve different specialists, a sometimes overlooked fact.

The design work for a survey, carried out by a team of methodologists and subject matter specialists, includes decisions about definitions of variables, geographical and other delimitation of the target population, questionnaire construction and testing, decisions about data collection methods, sampling design, sample size, nonresponse treatment and estimation method. It can contain elements of applied research. User requirements are respected as far as possible. Understanding users' needs and sorting out important from unimportant matters calls for skillful discussions with various parties involved, since it affects the proper allocation of funds. Equally important is the translation of the design into operational stages. It requires a great deal of experience to, for example, allocate cost in the best possible manner to the various operations that define the survey.

The process is a series of operations carried out by interviewers, coders, computer systems specialists, and data handlers who edit, impute, and make estimates. Some operations are carried out in the field (for example, data collection), others in the central office (for example, editing, coding, imputation, preparation of the data for estimation, and analysis). In a high-quality process, the participants function well together, under a project manager's competent supervision, in a reliable, well controlled manner.

Each operation and sub-operation affects data quality. Inadequate and poor training of the interviewers will increase the response errors. Inadequate interviewing manuals, poorly designed questionnaires, overly sensitive questions, and ambiguous definitions may give rise to measurement errors and nonresponse. Processing operations such as editing and imputation can, if uncontrolled, ruin the data quality. Maintaining high standards and full control in all these operations is costly, but even more costly to the organization may be the publication of data of inferior quality.

Most users are aware, remotely at least, of the difficulties that can arise in the process. They seek assurance, for example, through informative measures about vital aspects of the whole process. Design and process are to some degree independent activities, often handled by persons with different professional motivations. Specialization is necessary in a large organization. It can happen that a ''good'' (reliable and well-controlled) process results in statistics with unacceptably low accuracy (high MSE) because of a poor design. Here, the design is at fault, not the process. Conversely, an impeccable design may be followed by a poor process that ruins the data quality.

Do the statisticians succeed in informing the users adequately of the quality of the design and the process? Some aspects can be readily measured, such as rates for different causes of nonresponses, rates for edit and imputation, and so on. Other aspects are difficult to measure and communicate. For example, how to communicate evidence of insufficient training among the interviewers, or among some other category of participants?

5. The Idea of an Error Associated with Each Operation

Statisticians often mention that: (1) statistics production differs from other types of manufacturing in that errors are built in by design; (2) although statistical science dwells on deviations from a perceived ''true parameter value,'' one should avoid the negatively charged term ''error'' and speak instead in softer terms such as inaccuracy or deviation.

In regard to the first statement, it is true that, by design, observation is often limited, for cost reasons, to a relatively small sample of units, instead of the entire population, with a deliberate error as a result. But manufacturing of statistics is really not much different from, for example, manufacturing of automobiles, airplanes or refrigerators in the sense that production always involves a trade-off between cost and quality. In automobiles, the top of the line is expensive: the consumer pays more for Rolls Royce quality than for a more modest make. Likewise, in statistics, if additional resources can be spent in the production, better accuracy, higher quality will result. What may make statistics different is that a very serious quality deficiency, namely poor accuracy, is extremely difficult for the consumer (sometimes even for the producer) to perceive and complain about. For accuracy, the consumer/user is at the mercy of the professional competence of the producer/statistician.

As for the second comment, little can be done to alter the fact that statistical science is about probable deviations, inexactness, variability, and errors. Theoretical statisticians and methodologists are conditioned into thinking in terms of error. In a survey, if $\hat{Y}$ is the estimate (the published statistic) of the unknown quantity Y (the targeted population characteristic), the statistician's natural inclination is to analyze the error $\hat{Y} - Y$. By accuracy, he or she means a (usually probabilistic) statement about the magnitude of the squared error. Accuracy is a vital aspect of quality.

A survey consists of many operations. Early in the development of survey theory, it was found natural to associate an error with each operation. A ''perfect survey operation,'' one with zero error, is usually unattainable at a reasonable cost. Consequently, an operation with nonzero error is tolerated. Each error component influences the accuracy of the published statistics, and the total error, $\hat{Y} - Y$, is represented as the sum of the errors. The objective of total survey error modeling (see Section 8) is to account for at least the most relevant survey errors and to measure their relative importance.

Total accuracy evaluation means to evaluate all components of the total error $\hat{Y} - Y$. Accuracy in this total sense is of course what the user wants to know. Unfortunately, when the survey statistician talks about ''accuracy,'' he or she often refers to just one single error, the sampling error, measured by the sampling variance or the cv. The world of survey statisticians include a clan of ''sampling error specialists;'' many theoreticians and methodologists belong there. Their narrow focus is often criticized by other participants in the total process, not without reason. Further, the sampling error is often incorrectly measured by the sampling error specialists, because they do the computation of cv on data containing error, both the ''inescapable'' kind caused by response error and the ''self-inflicted'' kind created when the statisticians put artificial (imputed) values in place of missing real data.

Survey statisticians are rightly proud of 100 years of history containing some major developments and breakthroughs in the understanding of sampling error. The years 1897 (Kiaer's bold ideas) and 1934 (Neyman's convincing case for randomized selection) stand as important milestones.

Some say that the sampling error is ''over-researched'' at the expense of the other errors, which are claimed to be more important today because numerically greater than the sampling error in many surveys and because researchers neglected these other errors in the past. There is a hint that some sampling error specialists do little of practical value, turning, as they do sometimes, around mathematical fine points. Certainly, one can admit that many articles on the sampling error (new sampling schemes and new estimators) over these 100 years have had little impact on today's practice. But the intellectual level in the discussions has at times been inspiring. And despite what the antagonists say, sampling error theory has made advances in the last 30 years, with important results for large-scale computerized practice, such as the development of estimation systems (Bascula developed in The Netherlands, GES in Canada, CLAN in Sweden, CALMAR in France) applicable for general uses of auxiliary information.

Theoretically less developed was, for a long time, the edit operation, a normal part of survey processing aiming at completeness and logical consistency of the data set. Edit rules are used to identify wrong or ''suspect'' fields of a record. The identified fields are checked (by recontact, if possible) or may be replaced by imputed, possibly ''better'' values. Edit is costly if done at the micro level. Selective edit is a cost saving alternative which may not greatly increase the total survey error. Edit and imputation is now an important special field of professional competence. There is international co-operation among experts in the area, usually without much interaction with the sampling error specialists and their established mathematical tradition. To some theoreticians and many academics (especially those without close contact with a statistical agency) it is surprising that data are ''altered'' by seemingly arbitrary rules; to the uninitiated, this practice seems to violate the idea that data are sacred and not to be tampered with. Edit and imputation practices lie far from the material typically taught to statistics undergraduates, yet they are an inseparable part of statistics production.

A sign of vitality in the leading national statistical agencies is that they foster other narrowly focused but important competence. Examples are (i) Index theory and practice and (ii) National accounts theory and practice. For example, the Consumer price index, published in many countries by the national statistical agency, poses challenging issues with roots in statistical science, economics, and measurement theory. The National accounts experts form another important group in a statistical agency (whereas these experts seem to be rare in the Economics departments of the universities, despite the importance, for the nation, of the topic). For these specialists, high-quality workmanship has a basis in Economics and Statistical Science. They rely on accurate inputs in the form of data furnished by the survey statisticians.

6. Measuring and Reporting Design and Process Characteristics

The users want confidence in the quality of the statistics delivered to them. Confidence may be enhanced by providing users with objective measures of important aspects of the design and the process. A number of measures are in use. Some statistical agencies report some of the measures in some of their surveys. The policy, and the perceived usefulness, of such reporting is far from uniform. Too much numerical detail may be confusing to the user rather than enlightening.

Objective measurement of design and process is different from monitoring customer satisfaction with the statistical agency. A customer satisfaction index leaves out important aspects of quality, notably the accuracy of the results. Instead it focuses on service, punctuality, contents, ease of interpretation and other aspects that the customer can readily observe. The danger of pursuing a high score on a customer satisfaction index is evident. In order for the organization to feel good about itself, it will spend resources on those aspects that are readily noticed by the user, the external image, rather than on vital but less transparent aspects such as improved accuracy. Statistical agencies may feel an urge to shorten as much as possible the production time for the statistics they deliver: it is an obvious way to satisfy the user in an era that places a high premium on rapidity; it also increases the likelihood of gross errors and other inaccuracies in the statistics.

We believe that, at present, the primary source of user confidence is a feeling of general trust in the statistical agency, in the data it publishes and in its design and process work. This feeling is unmeasured, or unmeasurable, preconceived confidence, based on the reputation of the agency within the society it is serving, rather than on a number of objectively measured survey characteristics.

Trust in the national statistical agency varies considerably between countries; it is high in some countries, low in others. The statistical agency can succeed, over time, in building user confidence in their products. This takes sustained effort and investment in professional staff, and in advanced technology. The need is particularly strong in countries where the national statistical agency faces strong competition from private survey institutes. The users should be kept aware of the agency's investments in qualified personnel and improved technology. But there lies perhaps a danger in trust being equated with quality: the agency may become complacent, something which will not go unnoticed by the users for very long.

Particular user categories derive confidence from regularity of and familiarity with a given survey that they take an interest in. Continuous and periodic surveys have a reference to the past, so confidence may develop over time. One-time surveys must often stand on their own and may lead to uneasiness.

A number of the reported measures reflect imperfections in the process, elements which may cause harm to the accuracy. But the statisticians' message is incomplete. They fail, most of the time, to draw the full consequence of the reported imperfections; they do not inform the user on the real issue, namely the impact on accuracy.

To illustrate, a declared nonresponse rate of 36 percent is an objective figure but provides no answer to the real question: how is quality affected, how much greater is the MSE as a result? The statisticians are silent on this issue; the blame falls notably on the sampling error specialists. It is left up to the user to form his or her own idea about the harm done to quality. All users view high nonresponse one as of the main threats to data quality. But in assessing the effect of a stated rate of 36 percent, they risk falling back on stereotyped ideas, found perhaps in well-meaning texts, or heard from colleagues, of the type ''a nonresponse rate of more than 30 percent (or a similar rule of thumb) will render the survey data useless'' (and they may think this irrespective of the survey conditions). Lacking clearer advice from the statisticians, they can perhaps do little else but abide by such crude rules. The statisticians do not routinely measure the addition to the total MSE (a squared bias component, an additional variance component) that would unequivocally show the user how much harm is done to the accuracy as a result of 36 percent nonresponse.

Thus, the reporting of measures of survey characteristics raises questions such as these: What can the statisticians really tell about the impact on quality (notably on accuracy) of the various imperfections? At the receiver end, how is user confidence in the data altered by a wealth of performance figures and rates that may be received? There appears to be a gap between what the user is hoping for and what the statistician is capable of producing. From this angle, let us look at some quality indicators (other than the sampling variance).

6.1. Nonresponse patterns

Users view high nonresponse as a weakness in a survey, likely to cause biased estimates. Now the nonresponse consists of several components with different causes. Nonresponse component rates rank among the simplest measures that can be provided to users, including rates on refusals, no-one-at-home, eligibility, completion and so on. These rates may give the user valuable hints about the survey process, in particular about data collection. For example, a high refusal rate may suggest poorly trained and controlled interviewers. The total nonresponse rate (based on the sum of the components) is a crude measure. This is sometimes overlooked. When the statistician attempts to reduce nonresponse bias in the estimates by adjustment weighting within groups, the weights are usually based on the overall rate. This can be criticized as simplistic, because bias is the key question, and the different components of the total nonresponse may have quite different biasing effects.

6.2. Imputation

The goal of imputation is to ''complete the data set'' by inserting a ''reasonable'' value for one that is missing or cast in doubt by a failed edit. Many imputation procedures are in use. The imputed data set may be in some sense ''better'' than if imputation had not been attempted, as is expected when the imputed values are created with the aid of powerful auxiliary variables or by relevant historical information. A random hot deck does not improve the data but does create a complete rectangular data set, which is sometimes said to be the prime goal of imputation. But the key question is perhaps not so much the statistician's quest for increased information as the user's reaction to ''high'' rates of failed edit and imputation. The reporting of such rates is a normal part of quality declaration, but we know little about how the user's confidence in the data is modified by such information. High rates may generate some suspicion; ''reasonable'' rates are likely to pass unnoticed. Should the user not also be informed about the imputation method(s) in use? At any rate, for the statistician, imputation is a ''quick remedy,'' always ready to be used, as if nonresponse is nothing to worry about; imputation has entered the collective mind as a saving grace. The bias and the increased variance that it causes are seldom measured.

6.3. Reweighting

Reweighting of data is often used at the estimation stage to compensate for nonresponse. As in the example cited earlier, suppose the nonresponse rate in a survey is 36 percent and reported as such to the user, who may find it high. Now, the statistician carries out a reweighting of the data and may feel that this eliminates the high risk of quality deterioration (bias) that this rate represents. With skill and luck, he or she may be correct. Thus the high reported rate gives the user a very strong warning, but the reweighted data are reliable and fit for use. Should the user be given more detail about the reweighting? If based on adjustment groups, should the formation of the groups be declared, with a reminder of the accompanying assumption that all units (respondents and nonrespondents) in the same group have the same response probability? Does information about the reweighting method alter the user's confidence in the data, when he or she had already been told of the 36 percent nonresponse rate? Probably not.

6.4. Coverage

Despite efforts to maintain good lists, some units (households, persons, and business enterprises) are missed in a survey, although they are in the target population. Such coverage error has long been recognized as a problem, but its negative effects on quality (bias) are seldom or never considered. The issue becomes acute when coverage errors are particularly pronounced in certain population groups of interest. Statisticians as well as users know little about what may constitute an ''acceptable'' coverage error rate.

6.5. Administrative registers

In a number of countries, large quantities of statistics are derived from administrative sources, with or without some direct data collection. The accuracy of such statistics, and the user's confidence in them, depends on the quality of the administrative registers used. The statisticians lag behind in building a theory of accuracy assessment for such statistics. Needed is a theory that would gain public acceptance, similarly to the way that randomized sampling has become accepted.

In summary, we believe that it is useful for the agency to collect and keep available information on design and process characteristics such as those mentioned, for these reasons: (i) the information may find important uses within the agency, in that the indicators can identify weaknesses in data collection and data treatment and can suggest directions for research and development; (ii) the agency can, if called on, present the information to external users, as a testimony to its activities in regard to quality assurance. In addition, it may create an element of satisfaction for the agency to be able to refer to a measurement of important aspects of their surveys in words such as: ''We have the information on hand'' or ''We regularly disclose some information of this kind.'' Also, this activity may give the government agency an edge over the private survey institutes; the government agency may be better prepared than the private institutes to absorb the perhaps considerable cost of producing this information.

The question remains: what is the real value to the user of a host of detailed design and process information? More information than at present could be developed and provided. Where should the detail stop? Some users may out of habit monitor a few of the measures, but how do they process the rest of the information? We believe its total effect is limited; it probably does little to modify a preconceived confidence based on general trust in the agency and formed over a number of years of experience.

7. The Rise of Complex Issues for Research

The dominating base for survey theory research is randomization theory, or design-based theory. It has been convincingly defended by prominent spokesmen, such as Hansen, Madow and Tepping (1983). As Singh, Gambino and Mantel (1994) point out, ''Most producers of survey data are accustomed to design estimators and the corresponding design-based inferences. They interpret the data in the context of repeated samples selected using a given probability sampling design, and use estimated design-based cv's (coefficients of variation).''

The methods in the design-based paradigm require a known positive probability of observing any unit k and the data, say y_k, tied to this unit. This is satisfied if we sample from a perfect frame and there is no nonresponse. Otherwise, and that accounts for practically all surveys, the probability of observing a unit k is unknown. Consequently, randomization theory is strictly speaking incapable of dealing with ''normal'' surveys. The requirement of known positive probabilities puts the theory in a straitjacket; it cannot adapt. When the theory gained momentum in the 1930's, large-scale computerized edit and imputation was not yet practiced, and nonresponse was very low, so for some time the theory was not seriously challenged. Today, when these nuisance factors have reached levels far too important to be ignored, the design-based theory has to be ''fudged.'' This leaves many statisticians uneasy.

An urgent practical issue can give a decisive turn to the effort of the theoreticians. The practitioners ''drive'' the theoreticians, especially since there is a strong demand today on statisticians to deliver ''reasonable'' data despite clearly insufficient resources. In the beginning, an issue is given a simple, down-to-earth treatment by very practical people. Having been initiated and made aware, the theoreticians start employing their superior technical arsenal. More sophisticated, more technically advanced, more complete answers are offered. Some, or maybe much, of the results will, it is true, lie outside the immediate needs. The result is a stream of research involving considerable theoretical complexity. We mention two presently active research fields of this kind.

(i) **Insufficient accuracy of small area estimates.** Survey estimates for small domains (small areas) are, by definition, imprecise. A statistical agency may decide to suppress publication of a small domain estimate if the computed cv exceeds 30 percent. It is a wise policy. But some users, governments and others, demand estimates for domains so small that the agency's best design-based methods cannot deliver a decent accuracy. What should the statisticians do? A refusal to estimate does not seem a response worthy of a true professional. Hence users' demands, and the agency's willingness to oblige, open up rich research opportunities for the theoreticians. Many a published paper on small area estimation starts with a one-line apology ''There is a growing demand for reliable small area statistics from both public and private sectors'' and the following numerous pages of heavy theory are lost on the more practical statistician, who wonders if a miracle has occurred. No, it does not change anything. If only three values, say, are observed in a domain of interest, the accuracy of any estimate is unacceptably poor, as measured by the usual design-based cv. This holds whether the whole domains count 10 or 1,000 units. (Design is part of statistics; only a foolish designer would design a survey so that an area of interest is represented to the rate of only three observed values.) Even the best model dependent small area estimates will not change this fact. But science has progressed; much new knowledge has been produced. Some of it may open up new vistas. The survey manager is not fooled, knowing that for very small domains the accuracy remains unacceptably poor by accepted design-based standards.

(ii) **Measuring imputation variance.** Imputation is frequently carried out on a massive scale to compensate for survey nonresponse often in excess of 20 percent. It is expedient. It creates (i) an unknown bias in the estimates and (ii) additional variance over and above the sampling variance. Awareness of this has recently led researchers to seek new variance estimates that improve on older methods in that they include a component of imputation variance. Such variance estimates can be constructed in a variety of ''reasonable'' ways. Again, practice drives a proliferation of research and publications. Again, there is a flurry of research, and scientific progress, but in this example on an issue of rather minor importance, as compared to some of the ''neglected'' issues in surveys. The work misses somehow the central issue that imputation raises; this issue is bias in the estimates, not additional variance.

In these two examples, as in others that could be mentioned, the theoretician's delivery is of high technical caliber but remains incomplete. Editors of scientific journals may welcome the activity. Statisticians close to operations are not necessarily impressed. As Tortora (1996) remarks: ''...one hears from the research culture that their work is not (viewed as) important, is not used, or used with modification that negates much of its intended effect. And one often hears from the operations culture that the research is not pertinent to operational needs or is too complicated to work in practice.''

8. Creating a Functional Total Survey Error Model: An Impossible Dream?

A survey error model serves to conceptualize the errors associated with the various survey operations. The concept was developed in a series of important articles: Hansen, Hurwitz, Marks, and Mauldin (1951); Hansen, Hurwitz, and Bershad (1961); Hansen, Hurwitz, and Pritzker (1964). Known as the U.S. Bureau of the Census Survey Error Model, it has greatly influenced all subsequent thinking about survey error modeling. These authors analyzed a model with only two errors: sampling error and measurement error. But the foundation was laid for the more general concept of a total survey error model, one that would include a number of other relevant survey errors as well.

Hansen and his collaborators made innovations of lasting value when they created designs for sampling in two or more stages, making it possible to make unbiased estimates for households or individuals, despite the lack, in North America, of list frames. They soon realized that nonsampling errors are equally important or even more so. This led them to survey error models, whose objective is to measure the relative importance of each error and to make probability statements about the total error. But to date, the advances and the practical payoff of this admirable concept have been limited. Integrated modeling and joint estimation of survey error components is not an important research topic any more, although it should be.

Deciding what errors to include in an error model is the first task. Then we have to derive all the components of the total MSE arising from the model, squared bias terms, variance components, interactions of different kinds. The task quickly becomes unwieldy. We get a complicated set of expressions. If a practical purpose is to be served, we must then measure (estimate) all these components from available data. But in no national statistical agency that we know of is there a routine measurement of MSE components, not even for a fairly elementary model containing just a few error terms. At present it seems unrealizable to have a functional survey error model, one that would account for the major errors and yield estimates of the various components. For example, Platek and Gray (1983) analyze a survey model with three errors: sampling error, response measurement error, and imputation error (assuming that the nonresponse is treated by imputation). They arrive at intractable expressions for the components of the MSE. Simplifications would be necessary to obtain something practically useful.

Forsman (1989, 1993) reviews the history of survey error model theory, in particular its last 60 years. He notes that two lines of thought exist in regard to survey error models: (a) the development of theory for specific sources of nonsampling errors (recognizing that the sampling error has already experienced a rather long history), and (b) the development of an integrated treatment of survey errors. A positive development is some progress in regard to (a). It has brought increased awareness of the existence and the size of different nonsampling errors. Forsman notes, ‘‘. . . unfortunately, sub-fields in the subject (of survey modeling) tend to develop independently of each other, and often in ignorance of each other, despite the need for collaborative efforts. Although successfully integrated in the early development of survey models, the separate developments of survey research and planning of experiments in recent years illustrate the point.’’ We are far from realizing the integrated goal (b) of simultaneously measuring all important survey errors within one model. Lyberg, Biemer, and Japec (1998) express a similarly pessimistic view in regard to the future of total survey error modeling. Part of the difficulty is that experiments and other special arrangements may be required, before or during the survey, and in today’s tight budgetary conditions, statistical agencies seem unwilling to assume the cost for this.

Thus research on survey error is compartmentalized, rather than integrated. Researchers develop an expertise on a piece of the total error picture; collaborative effort is lacking.

The total error survey model of the 1950's has a tremendous appeal because it translates the complex sequence of the survey operations into a concise, analyzable mathematical statement. To measure the total error in an estimate is surely what statisticians ought to do. But now, 50 years later, the goal appears unrealizable. Leading statisticians express disappointment; as we noted, Smith (1994) feels that ''a systematic plan to reduce the total error is necessary if we are to have total quality management in surveys.'' Lacking such a systematic plan, backed by theory, survey managers instead follow simple rules of thumb for the survey operations. They are in the situation that Statistics Canada (1998) fittingly describes as having to ''make sensible choices'' within ''a collection of practices.'' Sensible choices are not necessarily scientifically based; perhaps they need not be?

It is not evident that a focus on customer satisfaction, with reference to a multifaceted quality concept, will facilitate progress on total survey error measurement. Instead, the emphasis shifts away from error measurement, by stressing instead user satisfaction on ''nonstatistical'' aspects of quality.

The two obvious reasons for wanting to persevere in the task of total survey error management are: (a) the relative importance of each error component needs to be measured and stated, in order to provide the user with objective information on the relative importance of the different errors, and to guide the survey statistician's decisions about the best use of additional resources, or the redistribution of existing resources, so that accuracy improvement is maximized; (b) a total accuracy measure is needed for any published statistic, so that an estimate of total MSE can be computed, not just (a dubious) quantification of the sampling variance. A confidence interval calculated on an estimate of sampling error alone may give the sampling error specialist 95 percent confidence, but may not give the user much confidence in the data.

The opposite view holds that total survey error modeling has had half a century to prove its mettle, and that routine measurement of total survey error now seems unlikely to ever happen. A compartmentalized line of attack, on individual survey errors, may subsist. Besides, all of the statisticians in the agency do not share the urge of the theoreticians and the methodologists to measure all important errors. For example, those trained in psychometrics are more likely to focus on reliability, validity and other aspects of questionnaires; to them decomposition of the total survey error does not seem a priority.

9. Is Survey Theory and Practice in a Crisis?

Why does national statistics production appear to be a collection of practices without a strong theory? Statistics Canada's opinion that survey methodology is ''a collection of practices, backed by some theory and empirical evaluation'' is probably shared by other national statistical agencies. ''Hard sciences'' have firm and reasonably complete theory; physics is an example that we envy. But even artistic fields such as literature and music claim to have theory, or theories. Why not the production of statistics?

Classical models of scientific progress, embraced by theorists of science such as Popper and Reichenbach, hold that it is only possible to speak of scientific progress if knowledge is acquired through purely cumulative theories. Cumulative knowledge is seen as the ideal. Progress in survey theory has not advanced in that fashion: (i) Randomization theory (design-based theory) has resisted progress beyond the imaginary world devoid of nonresponse and faulty frames, because with these nuisance factors we no longer have known positive probabilities of observing the data y_k tied to unit k; (ii) Total survey error theory has resisted progress partly because of the unknown probabilities, partly because of complex MSE components (squared bias terms, variances, interactions) and the high cost of measuring these components.

Randomization theory and Total survey error theory have not been successful theories. This does not mean that they are useless theories. They have brought guidance and support for practice, on a number of questions of limited scope, and the 20th century witnessed some celebrated breakthroughs. But they have not been particularly successful in achieving a satisfactory cumulative progress, something many survey statisticians had hopes of.

Margenau (1961) offers a view on the progress of theory as a cumulative, expanding process; it was a typical view at least before the postmodern era: ''The process of scientific explanation moves through successive steps from the particular to the universal. A given fact, or set of facts, is first interpreted by means of a fairly simple model. The first model, however, is likely to be limited in its range of application. Science can develop in two ways: it can employ other, different simple models for other parts of the evidence, resigning itself to the use of incompatible explanations for disparate experiences on the motto of ''different theories for different observations.'' The most thoughtful scientists, however, reject multiplicity of models, one for each domain of science; they search for a system of concepts extensible to as large as possible a domain of facts.''

This explanation of progress has relevance for survey theory. The sampling error theory of the 1930's relied on ''a fairly simple model,'' that of randomized sample selection with known inclusion probabilities from a completely listed finite population. As time went by and practice evolved, that theory had to come face to face with such nuisances as high nonresponse, and automated edit and imputation for large masses of data. The simple randomization theory did not suffice. Now ''thoughtful scientists'' seek a broad, expanding, cumulative system, not one model for each survey operation. They want to cover as wide a domain of facts as possible in one complete theory. It is in this spirit that Smith (1990, 1994) deplores the lack of a systematic total survey error theory and measurement. The concept of total survey error modeling in the 1950's was meant to expand survey theory, but it did not happen.

Doubts are expressed, occasionally, about the directions of survey theory. For example, Hinkins, Oh, and Scheuren (1997) deplore the awkward situations that arise when we try to extend the too narrow base offered by randomization theory. They deplore ''all the fiddling we have to do when trying to correct for nonsampling errors.'' The more ''fiddling'' there is, the more the theoretical base will seem insufficient.

The randomization theory of Neyman, Hansen and their contemporaries was operating in a near perfect world of negligible nonresponse rates, known probabilities of observing the units, and few other nonsampling errors. Today, we have to ''fiddle'' with ''corrections'' which entail increasingly complex maneuvers and which result in a limping theory. At the same time, this satisfies the theoretician's desire to supply ingenious ''solutions'' to limited problems. Among the practical statisticians, further doubt is cast on getting firm guidance from ''broad solid theory.''

Kish (1995) reviews one hundred years of survey theory development. He comments on how cluster sampling, which existed as a practice already around 1910, took a long time to come under the umbrella of proper theory, and he notes that ''theory lagged behind practice then and still does.'' For a number of practices in our present ''collection of practices,'' there is indeed a strong need for theory to catch up. Further, Kish (1995) recommends that the calculation of sampling error (design-based cv) should not be abandoned, despite the limited light it sheds on the total error picture. He asks: ''Are sampling errors necessary and sufficient, when other survey errors, such as measurement, are often potentially larger but unknown?'' His answer is important, in part for its allusion to small area estimation: ''Yes, they are necessary, though not sufficient. And they become relatively more important for small subclasses and for comparisons and other analytical statistics.'' Sampling error measurement is one element of survey methodology, one of the few that rests on solid theory foundation. Sampling has achieved public acceptance. It is upheld by an important scientific stream. Capable scholars have been able to lift some of the mundane practical questions arising in surveys onto a scientific stage. Fundamental principles of survey sampling and estimation are discussed, as recently in Brewer (1999).

It remains to be seen if the present theory base will endure or if a scientific revolution will take place. What would replace the probability based notions of accuracy that the present generation of statisticians learned and extended, on the foundations from Neyman, Hansen, and others? In the future, will ever more powerful computers somehow succeed in making better sense out of large survey data sets, appealing perhaps to radically different notions of accuracy and ''quality in statistics?''

10. Conclusion

Our title asked: Can a statistician deliver? Our main objective has been to pose the question and to offer some thoughts around it. We have no complete answer. Most national statistical agencies strive hard to provide high quality statistics. In this effort they rely on the contributions of individual, well trained statisticians of different kinds. In today's complex survey environment, one can point to factors which make it difficult for an individual statistician to deliver in a complete sense. In this article we have noted some of these circumstances:

1. Quality, as officially defined (we cited two such definitions) is a multidimensional concept. It is difficult for an individual statistician to feel motivated in regard to all of its many components.
2. Some argue that survey methodology is a set of practices with no unifying theory covering the entire process. The statisticians would be helped if there were a more complete and coherent theory. It would make them (in particular the methodologists and the subject matter statisticians) feel more justified and secure in their daily practice. A complete theory is not in sight.
3. The user wants confidence both in the survey design and in the survey process. The statistical agency can make it a rule to publish numerical information on various

indicators. But some measures that might be important for building user confidence may be sensitive and difficult to report. In any case, information on a large number of design and process features is unlikely to significantly change user confidence, which is based largely on a general, preconceived trust in the agency. Moreover, the statisticians do not deliver a clear message to the users on the exact consequences (the loss of accuracy) that survey imperfections may have. But producing survey performance measures is important for the organization itself.

4. There is some reluctance among theoreticians to focus their research on other questions than those for which an answer amounts to a probability based assessment of error, in particular the sampling error. Variance is overemphasized; too little is known about different types of bias.
5. Total survey error modeling is an admirable concept for systematic evaluation of total accuracy, a crucial dimension of quality. But the statisticians have not been able to deliver on the promise that this concept once held. On the other hand, the study of specific error components, including several of the nonsampling errors, has made some progress.

11. References

Brewer, K.R.W. (1999). Design-based or Prediction-based Inference? Stratified Random vs Stratified Balanced Sampling. International Statistical Review, 67, 35–47.

Colledge, M. (1996). Comment on ‘‘Why Innovation is Difficult in Government Surveys’’ by D.A. Dillman. Journal of Official Statistics, 12, 145–150.

Dillman, D. (1996). Why Innovation is Difficult in Government Surveys. Journal of Official Statistics, 12, 113–124.

Elvers, E. and Rosén, B. (1998). Quality Concept for Official Statistics. Encyclopedia of Statistical Sciences. New York: Wiley.

Forsman, G. (1989). Early Survey Models and Their Use in Survey Quality Work. Journal of Official Statistics, 5, 41–55.

Forsman, G. (1993). Recent Advances in Survey Error Modeling. Jahrbücher für Nationalökonomie und Statistik, 211: 3–4, 331–350.

Hansen, M.H., Hurwitz, W.N., Marks, E.S., and Mauldin, W.P. (1951). Response Errors in Surveys. Journal of the American Statistical Association, 46, 147–190.

Hansen, M.H., Hurwitz, W.N., and Bershad, M.A. (1961). Measurement Errors in Censuses and Surveys. Bulletin of the International Statistical Institute, 38, 2, 359–374.

Hansen, M.H., Hurwitz, W.N., and Pritzker, L. (1965). The Estimation and Interpretation of Gross Errors and the Simple Response Variance. In Contributions to Statistics Presented to Professor P.C. Mahalanobis on the Occasion of his 70th Birthday. Oxford: Pergamon, 111–136.

Hansen, M.H., Madow, W.S., and Tepping, B.J. (1983). An Evaluation of Model Dependent and Probability Sampling Inferences in Sample Surveys. Journal of the American Statistical Association, 78, 776–807.

Hinkins, S., Oh, H.L., and Scheuren, F. (1997). Inverse-Sampling Design Algorithms. Survey Methodology, 23, 11–21.

Kish, L. (1995). The Hundred Years' Wars of Survey Sampling. Statistics in Transition, 2, 813–830.

Lyberg, L., Biemer, P., and Japec, L. (1998). Quality Improvement in Surveys – A Process Perspective. Joint Statistical Meetings, American Statistical Association, Dallas, August.

Margenau, H. (1961). Open Vistas. Philosophical Perspectives of Modern Science. New Haven and London: Yale University Press.

Morganstein, D. and Marker, D. (1997). Continuous Quality Improvement in Statistical Organizations. In Survey Measurement and Process Quality, eds. L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin, New York: Wiley, 475–500.

Platek, R. and Gray, G. B. (1983). Imputation Methodology: Total Survey Error. In Incomplete Data in Sample Surveys, eds. W.G. Madow, I. Olkin, and D. Rubin, Vol. 2, New York: Academic Press, 249–333.

Singh, M.P., Gambino, J., and Mantel, H.J. (1994). Issues and Strategies for Small Area Data. Survey Methodology, 20, 3–22.

Smith, T.M.F. (1990). Comment on Rao and Bellhouse: Foundations of Survey Based Estimation and Analysis. Survey Methodology, 16, 26–29.

Smith, T.M.F. (1994). Sample Surveys 1975–1990; An Age of Reconciliation? First Morris Hansen lecture; with discussion. International Statistical Review, 62, 5–34.

Statistics Sweden (1994). Kvalitetsbegrepp och riktlinjer för kvalitetsdeklaration av officiell statistik. (Quality Concepts and Recommendations for Quality Declarations of Official Statistics). Statistics Sweden MIS 1994: 3. (In Swedish).

Statistics Canada (1998). Statistics Canada Quality Guidelines. Third Edition. October.

Tortora, R.D. (1996). Comment on ''Why Innovation is Difficult in Government Surveys'' by D.A. Dillman. Journal of Official Statistics, 12, 179–183.

Received April 2000