Journal of Official Statistics, Vol. 19, No. 4, 2003, pp. 343–363

Reporting Sources of Error in U.S. Federal Government Surveys

Daniel Kasprzyk[1] *and Lee Giesbrecht*[2]

In 1996 the United States Federal Committee on Statistical Methodology (FCSM) established a subcommittee to review the measurement and reporting of data quality in federal data collection programs. The issues contained within this broad mandate are complex. Data user goals, and consequently their concept of ''quality,'' are varied. Similarly, reporting about quality can be implemented in various ways depending on the data product.

The FCSM subcommittee, whose membership represents the experiences of twelve U.S. statistical agencies, approached this topic by focusing on the accuracy dimension of quality and asking: 1) What measurement methods do federal data collection programs use to assess sources of error?; 2) To what extent do federal data collection programs report information on sources of error to the user community?; 3) How does reporting about error sources vary across different types of publications and dissemination media?; and 4) What information on sources of error should federal data collection programs provide and how should they provide it? The subcommittee completed three studies that focused on reporting sources of error in each of three types of data dissemination products; short-format reports, analytic reports, and the Internet. The studies provided information for the subcommittee about U.S. government statistical agencies' practices with regard to reporting the accuracy of their data (McMillen and Brady 1999; Atkinson, Schwanz, and Sieber 1999; Giesbrecht, Miller, Moriarity, and Ware-Martin 1999).

The subcommittee completed a report that discusses quality in terms of the measurement and reporting of various error sources that affect data quality: sampling error, nonresponse error, coverage error, measurement error, and processing error. The report discusses the measurement of each source of error – the measurement techniques and methods used; then it presents current practices for reporting information about the error source; and, finally, it presents recommendations for measuring and reporting survey error. This article summarizes the results of the studies of current reporting practices and provides recommendations to improve reporting of information on the various sources of error.

Key words: Survey error sources; nonsampling; measurement; coverage; processing; nonresponse; sampling.

[1] Mathematica Policy Research Inc., 600 Maryland Avenue, SW, Suite 550, Washington, DC 20024-2512, U.S.A. Email: DKasprzyk@Mathematica-MPR.com

[2] U.S. Bureau of Transportation Statistics, 400 Seventh Street, SW, Washington, DC 20590, U.S.A. Email: Lee.Giesbrecht@bts.gov

Acknowledgment: The authors would like to acknowledge the members of the Subcommittee on Measuring and Reporting the Quality of Survey Data for their efforts in conducting the research and developing the recommendations presented here. The subcommittee members were Daniel Kasprzyk (Chair), Dale Atkinson, Judy Conn, Ram Chakrabarty, Charles Darby, Brian Harris-Kojetin, Howard Hogan, Nancy Kirkendall, Lee Giesbrecht, Marilyn McMillen Seastrom, Renee Miller, Chris Moriarity, Dennis Schwanz, Carolyn Shettle, W. Karl Sieber, Antoinette Ware-Martin, John Wolken, and Graham Kalton (Senior Advisor). This article summarizes work undertaken by U.S. federal government staff as members of the Subcommittee on Measuring and Reporting the Quality of Survey Data of the U.S. Federal Committee on Statistical Methodology. The subcommittee's full report is published as Statistical Policy Working Paper Number 31, *Measuring and Reporting Sources of Error in Surveys* (Washington, DC: U.S. Office of Management and Budget).

1. Introduction

The U.S. Office of Management and Budget (OMB) coordinates the work of over 70 statistical agencies, 8 of which are independent and the remaining 62 of which are spread among 10 separate departments. In 1996, the OMB's Federal Committee on Statistical Methodology (FCSM) established a subcommittee to review the measurement and reporting of data quality in data collection programs of this decentralized statistical system. The subcommittee's mandate was challenging because of the number of federal agencies the OMB works with, the number of users of federal data at different levels of analytic skills and with different analytic goals – and, consequently, myriad ideas of what constitutes ''quality.'' In addition to multiple users and multiple concepts of quality, the measures of quality reported vary depending on the data product the data collection organization prepares and releases to analysts and the public.

A substantial literature exists and continues to grow on the topic of survey data quality (Lyberg, Biemer, Collins, de Leeuw, Dippo, Schwarz, and Trewin 1997; Collins and Sykes 1999; Dippo 1997; Carson and Liuksila 2001) and its management in national statistical agencies (Brackstone 1999; Colledge and March 1997). A number of conferences have taken place over the last ten to fifteen years related to data quality in National Statistical Offices. See, for example, Statistics Canada (1991, 2002), Statistics Sweden (2001a), the U.S. Federal Committee on Statistical Methodology (1991), and the International Conference on Survey Measurement and Process Quality (Lyberg et al. 1997). Invited sessions on the topic of survey data quality have occurred regularly at U.S. Federal Committee on Statistical Methodology sponsored conferences (U.S. Federal Committee on Statistical Methodology 1997, 1999, 2000, and 2001b).

Following the model of several authors, key concepts of quality include the following: accuracy, relevance, timeliness, and accessibility. Unanimity concerning the dimensions of quality does not exist. Statistics Canada, for example, identified interpretability and coherence as two additional dimensions, resulting in six dimensions of quality (Brackstone 1999). Statistics Sweden identifies five dimensions of quality (Statistics Sweden 2001b). The European Statistical System identifies seven dimensions, the three additional dimensions being comparability, coherence, and completeness (Eurostat 2000a, 2000b). While the number of dimensions related to the concept of quality varies, the central idea underlying it is that the characteristics of the data product in preparation meet or exceed the stated or implied needs of the user. The dimensions of quality are discussed below.

Accuracy is an important and visible aspect of quality that statisticians and survey methodologists have been concerned with for many years. Accuracy relates to the distance between estimated and true (unknown) values. For many, accuracy means the measurement and reporting of estimates of sampling error in a sample survey, but in fact the concept is much broader, taking in nonsampling error as well. Nonsampling error includes coverage error, measurement error, nonresponse error, and processing error.

Relevance refers to the idea that the data collection program measures concepts that are meaningful and useful to data users. The concept is critical to obtaining useful data because of its relationship to the question of whether the survey concepts implemented in a data collection program fit the intended uses. Determining the relevance of concepts

and definitions is difficult and time-consuming, requiring the expertise of data collectors, data providers, data users, agency researchers, and expert panels.

Timeliness can refer to two concepts. First, it can refer to the length of the data collection's production time – the time from the end of data collection until the first availability of a product. Without exception, end users prefer fast release times. Second, it can refer to the frequency of the data collection, that is, how often data are collected and reported.

Accessibility, as a characteristic of data quality, refers to the ability of data users to obtain the products of the data collection program. Data products are most valuable when they are easily available to end-users and in the forms and formats desired. Accessibility also implies that the data products include adequate documentation and discussion to allow proper interpretation of the results.

Other dimensions mentioned earlier identify four other characteristics of data quality: interpretability of statistics, comparability of statistics, coherence, and completeness. In this context, interpretability of statistics refers to the availability of information to help a data user understand statistics produced by National Statistical Organizations; comparability of statistics refers to the ability to make reliable comparisons over time; coherence refers to the ability of the statistical data program to maintain common definitions, classifications, and methodological standards when data originate from several sources (Brackstone 1999); completeness refers to the ability of the statistical data collection to provide statistics for all domains identified by the user community (Lyberg et al. 2001).

The subcommittee limited its coverage to discussing the accuracy dimension, a dimension that has a history of measurement and reporting, and interpretability, a dimension of quality that relates to the reporting of accuracy. Statistical indicators that describe different aspects of survey accuracy were reviewed in relation to various error sources, how indicators are measured, and whether and how they are presented to data users. The following four questions helped guide the approach taken:

- What measurement methods do federal agencies use to identify sources of error in data collection programs?
- To what extent do federal data collection programs report information on sources of error to the user community?
- How does reporting about error sources vary across different types of publications and dissemination media?
- What information on error sources should federal data collection programs provide and how should they provide it?

''Data quality'' was reviewed in terms of the measurement and reporting of error sources following the typology described by Kish (1965): sampling error, nonresponse error, coverage error, measurement error, and processing error. Others, in particular, the European Statistical System, however, have noted the importance of another error source, model assumption errors (Davies and Smith 1998). Although the importance of model assumptions is recognized, particularly in the analysis of business survey data, the Kish typology was followed in discussions of sources of error in surveys. Although the accuracy dimension is only one aspect of a multidimensional concept, even this dimension had not been systematically studied across the U.S. statistical system. The report from which this

article follows provides a chapter on each error source and includes recommendations for the types of information about each error source that should accompany short reports, analytic reports, technical or methodological reports, and be available on the Internet. The report concludes with a chapter that discusses several approaches to the measurement and reporting of total survey error and provides several recommendations for National Statistical Offices.

An understanding of current practices in reporting the quality of survey data is largely dependent on anecdotal evidence and the experiences of individuals working within agencies and survey programs. Three studies were conducted to characterize current agency practices in reporting sources of error. The studies dealt with three reporting formats: the short-format report, the analytic report, and the use of the Internet. The results of these studies provided a framework for a discussion of issues and recommendations. This article summarizes the analyses of current reporting practices (Kasprzyk et al. 1999) and the recommendations for improving reporting of information on the various sources of error (U.S. Federal Committee on Statistical Methodology 2001a).

2. Report Formats and Reporting Sources of Error

Report formats for presenting statistical results to users vary considerably not only across statistical agencies but also within a single statistical agency. Although most agencies and sub-agencies recognize the importance of reporting information about the survey design and the nature and magnitude of survey error sources, they do not always agree about how much detail to provide in the different reporting formats.

Generally speaking, statistical agencies agree that basic information about the survey's purpose, key variables, sample design, data collection methods, and estimation procedures ought to be available in descriptive and analytic reports. Additionally, there is a consensus that reports of results should also describe and account for sampling, nonresponse, coverage, measurement, and processing error. But, while there is general agreement that this information should be reported, there is no clear answer as to how much information to provide in the various reporting formats. A long, reasonably detailed discussion of error sources at the end of a lengthy complicated analytic report may seem reasonable in that context. Such detail is obviously inappropriate for a short 2–10 page report, but striking a reasonable reporting balance can be difficult because there is a strong belief that some information on the data source and error sources should be reported, regardless of the length of the report.

Obviously, details reported ought to depend on the nature of the report and its intended use. Individual data programs have a variety of constituencies and user groups, each with diverse representation – ranging from sophisticated data analysts with graduate degrees to reporters and the general public. These user groups are served by different types of data products, which the subcommittee addressed in three broad categories; short-format reports, analytic reports and accompanying technical reports, and the Internet.

3. Reporting Error Sources in Short-Format Reports

Short reports, directed to specific audiences such as policy-makers and the public, focus typically on a narrowly defined topic, issue, or statistic. At the outset, subcommittee

members varied on their opinions of the level and amount of technical detail to include in such reports. The subcommittee began this work with an assessment of current practices.

3.1. Reporting error sources in short-format reports: results of a study

McMillen and Brady (1999) reviewed 454 publications of ten pages or less to examine the treatment of information about survey design, data quality, and measurement. The publications are products of the 12 statistical agencies that constitute the U.S. Interagency Council on Statistical Policy and are available over the Internet. The majority of the reports were published in the mid-1990s or later.

The study found considerable variation in the amount of technical documentation included across this reporting format. Since there is little consensus about the amount of detail to be included in short reports, the authors limited the review of the reports to noting whether or not the reports *mentioned* specific error sources or specific elements of documentation about survey design.

Virtually all of the short reports that were reviewed included some information on how to learn more about the data reported. This information ranged from a contact name, telephone number, and e-mail and Web site addresses to citations of printed reports. Approximately two-thirds (69 percent) of the 454 reports included either a reference to a technical report or some mention of study design, data quality, or survey error. Close to one-half (47 percent) included some information describing the purpose of the survey or analysis, the key variables, the survey frame, and/or key aspects of the sample design. Only 20 percent of the publications included the sample size and 10 percent described the mode of data collection. Only a very small fraction (2 percent) mentioned estimation and/or weighting.

About one-fifth (22 percent) mentioned sampling error. In most cases, this was no more than a mention, although occasionally a report would refer to statistical significance testing and significance level. Only a handful of reports included information on the size of the sampling error. Nonresponse error is the most visible and well-known source of nonsampling error and certainly the most recognizable indicator of data quality, but only 13 percent of the short reports included any reference to response rates, to nonresponse as a potential source of error, or to the use of imputation methods to compensate for item nonresponse. Only 3 percent reported unit nonresponse rates, and there was virtually no reporting of item nonresponse rates. Approximately 10 percent of the reports covered mentioned coverage rates or coverage as a potential source of error. The difficulties associated with potential sources of measurement error were reported in 22 percent of the reports reviewed. Processing errors as a potential source of survey error were cited in 16 percent of the reports.

These results are not surprising. The reports studied are short and oriented to a specific topic; the principal goal of the publication is to convey important policy-relevant results with a minimum of text. However, the disparity between stated policy and implemented policy concerning the reporting of error sources is obvious.

3.2. Reporting error sources in short-format reports: discussion and recommendations

The short-format report limits the amount of data quality information that can be presented. Nevertheless, the essential principle of reporting information on the nature and

magnitude of error must be addressed. The subcommittee recommends that:

> *All short-format reports shall provide basic information about the data set used in the analysis, known sources of error, related methodological reports, and a contact for further information.*

The information presented must be brief, yet sufficient for the reader to appreciate the limitations of the methodology and data. Thus, the report should include the name and year of the data collection program the analyses are based on, and whether the data are based on a probability sample or census. It should also state that the data reported are subject to sampling error (if a sample survey) and nonsampling error. The total in-scope sample size and the overall unit response rate should be reported. Reports that describe findings should state whether statistical significance testing was used, and reference the significance level. All reports should include a statement that sampling errors for estimates in the reports are available on request. When only a few estimates are displayed, presenting confidence intervals associated with the estimates may be appropriate. Estimates dependent on survey variables with high item nonresponse rates should be identified, as should those with particularly difficult measurement properties. A report that includes more detailed information about data collection and data quality should be cited, along with the name of a contact person who can provide additional information or answer questions.

The information in the above recommendation may be conveyed in a short paragraph at the conclusion of a short-format report. The subcommittee recommends that agencies adopt a reporting format that can be repeated with only minor modifications across their short-format reports. One example that could be used for short reports that are a few pages in length might look like this:

> *Estimates in this report are based on a national probability sample of <Sample Size> drawn from the <Sampling Frame>. All estimates are subject to sampling error, as well as nonsampling error, such as measurement error, nonresponse error, data processing error, and coverage error. Quality control and editing procedures are used to reduce errors made by respondents, coders, and interviewers. Statistical adjustments have been made for unit nonresponse and questionnaire item nonresponse. All differences reported are statistically significant at the 0.xx level. The response rate for the survey was xx.x percent. Sampling errors for the estimates in this report are available from <Sampling Statistician (phone number; e-mail address)>. Detailed information concerning the data collection (including procedures taken to test the wording of questions), methodology, and data quality are available in <Data Collection and Methodology Report>. For more information about the data and the analysis contact <Program Contact (phone number; e-mail address)>.*

4. Reporting Error Sources in Analytic Reports

A second study conducted by members of the FCSM Subcommittee dealt with ''analytic publications.'' These publications include a wide variety of reports, report series, and types of analyses covering a wide range of topics. The publications use a variety of formats, with results described in narrative form, displayed in tables, shown in graphical format, or a combination of these. Analytic publications, as we considered them, may describe analyses of specific hypotheses or may summarize data from one-time surveys

or an ongoing series of surveys, where the surveys may be of individuals, households, establishments, or farms, etc. The most important characteristic of this type of report is that it provides fairly detailed analyses and/or summaries of data and that its length is not constrained to a few pages. By design, the reports themselves are longer than the short-format reports described previously, and give more opportunity for analysts to research a topic and provide analyses. As a result, data providers, survey methodologists, and statisticians also have more opportunity to describe the sources and limitations of the data.

4.1. Reporting error sources in analytic reports: results of a study

The second study, undertaken by Atkinson et al. (1999), reviewed 49 analytic publications produced by 17 agencies. The review included publications from major statistical agencies as well as some from smaller agencies conducting surveys. The selected publications were a convenience sample, but an effort was made to cover as many of the major statistical agencies as time would allow. The publications reviewed and the criteria and procedures used for evaluating the presentation of information on sources of error are found in Atkinson et al. (1999). They developed two sets of criteria, one focusing on the reporting of error sources, the other on the completeness of background information describing the survey. To standardize and document the review of the publications, they defined hierarchical levels of increasing detail for each of the major categories of the review criteria, recognizing the subjective nature of the classification and levels of detail.

The study considered both the completeness of background information on survey design and procedures and reports of error sources. Sampling error was the most frequently documented error source, being mentioned in 92 percent of the reports. Among the analytic reports reviewed, 75 percent presented sampling errors, 75 percent gave a definition and interpretation, and 45 percent specified the method used in calculating sampling errors. Somewhat surprisingly, only 71 percent mentioned unit nonresponse, 59 percent reported an overall response rate, and 20 percent reported response rates for subgroups. Only one-half (49 percent) mentioned item nonresponse and only 22 percent reported any item response rates. Nearly all reports included a definition of the universe (94 percent) and identified and described the frame (84 percent). But only one-half (49 percent) specifically mentioned coverage error as a potential source of nonsampling error, and only 16 percent provided an estimated coverage rate. Two-thirds of the reports mentioned measurement error and one-half included a description and definition. Specific studies to quantify this error were mentioned in only 18 percent of the reports. The majority of the reports (78 percent) mentioned processing as an error source, but very few included any detail about this error source (about 4 percent reported coding error rates and 6 percent reported edit failure rates).

The study indicated that survey background information was reported reasonably well – the general features of the sample design were reported about 92 percent of the time, data collection methods were reported about 88 percent of the time, and a brief description of the estimation techniques was given about 82 percent of the time. The review of error sources revealed variation across agencies. Only 59 percent of the reports included at least some mention of each of the five error sources.

The results of this study are not comforting. While recognizing that the evaluation criteria were somewhat subjective and granted the obvious limitations of a small convenience sample, the fact remains that a considerable difference exists between stated principles of practice and their implementation when it comes to reporting sources of survey error.

4.2. Reporting error sources in analytic reports: discussion and recommendations

For this study, analytic reports include a very broad group of reports and analyses that are not easily characterized. Many government reports can easily fall into this category, a category that obviously excludes short-format reports of a few pages as well as press releases. Also excluded from this category are compendia reports or statistical digests, that is, reports that bring together data from many data sources. The most important characteristic of the analytic report is that it provides fairly detailed analyses and/or summaries of data from either one-time or continuing surveys. Because of their length, analytic reports give more opportunity for data providers and analysts to describe the sources and limitations of the data. The subcommittee's recommendations acknowledge this opportunity while recognizing that the purpose of the report is to present statistical information and analyses. The length limitations of the short-format reports do not apply and, consequently, authors may more fully describe the survey methodology and data limitations. On the other hand, the treatment of methodology in an analytic report cannot be so lengthy and detailed that this information overshadows the statistical information presented.

Analytic reports are usually intended for a broad and multi-disciplinary audience. The reports usually provide technical notes or a methodology appendix that contains information about the data sources and their limitations. A critical aspect of the recommendations is the understanding that such technical information must provide the essentials: the key aspects of the survey background, and the major sources of error in the survey. The details of the data collection operations and procedures, studies about the error sources, and detailed analyses of the effects of statistical and procedural decisions belong separately in their own technical reports, comprehensive design and methodology reports, or quality profiles. The technical appendix does not need to be lengthy – it should be on the order of 5–10 pages – but it should provide quantitative information about the sources of survey error as well as citations of secondary sources that provide more detailed information or analyses.

The Subcommittee recommends that:

> *Studies reporting analyses of statistical data should present three types of information:*
>
> 1. *Background description of the data collection programs used in the analysis; the table below lists key information.*
> 2. *Description of each major source of error, the magnitude of the error source (if available), and any known limitations of the data.*
> 3. *Access to the questionnaire or questionnaire items used in the analysis, through the report, on request, or through electronic means.*

4.3. Reporting error sources in analytic reports: survey information

Table 1 identifies topics related to the background of the survey that ought to be addressed in a technical appendix. These are topics important with regard to satisfying

Table 1. Analytic report: recommended background survey information

Survey objectives	Data processing
Survey content	• Identification of procedures used to minimize processing errors
Changes in content, procedures, and design from previous rounds	• Editing operations
Survey preparations/pretests	• Coding operations
	• Imputation methods
Sample design	
• Target population defined	Estimation
• Sampling frame identified and described	• Description of procedure (stages of estimation)
• Stratification variables	• Source and use of independent controls
• Sample size	
	Key variables/concerns defined
Data collection	
• Schedule (when collected/number of follow-ups/time in field)	
• Mode (percent of each type)	
• Respondent (identified/percent self/ percent proxy)	
• Reference period identified	
• Interview length	

Recommendation 1 above. The discussion below concerns individual sources of survey error and recommendations related to the reporting of each source of error. The specific recommendations for each error source are the components to successfully implementing Recommendation 2. Recommendation 3 is obvious, but, unfortunately, is not always followed. The questionnaire items used in analysis, or the questionnaire itself, perhaps electronically or on request is important for analysts. The availability of the questionnaire allows analysts to understand the context of the question addressed to the respondent. Electronic questionnaires used in Computer Assisted Personal Interviews (CAPI) and Computer Assisted Telephone Interviews (CATI) ought to be made available to users in a way that preserves the original format as much as possible, for example by providing pictures of computer screens as presented to the interviewers along with the accompanying skip logic.

The information identified in the recommendations helps readers/users of analytic reports to improve their understanding of a report's findings. A substantial amount of information described in the recommendations is typically available in the sampling, data collection, data processing, and statistical estimation specifications. The difficult task for the data producer is to synthesize and compress the available material into an informative technical appendix.

4.4. Reporting error sources in analytic reports: accuracy of estimates

The information in Recommendation 2 concerns reporting the accuracy of the estimates. All statistical agencies address this issue in some fashion, but the information is presented inconsistently. Basic statistical data to inform users of the quality of the data collection operations are often not reported. Substantial gaps exist in the reporting of quantitative

information. Following the error typology of Kish (1965) described earlier, specific recommendations follow concerning the reporting of sampling, nonresponse, coverage, measurement, and processing error in analytic reports, technical documentation or user guides.

4.4.1. Reporting error sources in analytic reports: sampling error

Sampling error is probably the best-known source of survey error, and refers to the variability that occurs by chance because a sample rather than an entire population was surveyed. Gonzales, Ogus, Shapiro, and Tepping (1975) provide an excellent discussion concerning the presentation of sampling errors in surveys. For any survey based on a probability sample, data from the survey can be used to estimate the standard errors of survey estimates. Nowadays, the standard errors for most estimates can be readily computed using software that takes account of the survey's complex sample design. The challenge that occurs with the computation of standard errors occurs as a result of the multi-purpose nature of many federal surveys. Surveys produce many complex statistics and the task of computing and reporting standard errors for all the survey estimates and for differences between estimates is an extremely large one. The following applies regarding information about sampling error in analytic reports:

- Users must know if data are from a random sample or nonrandom sample; if the latter, then the implications for inference should be described.
- Sampling error should be identified as a source of error; it should be explained and interpreted for data users.
- If statistical tests are used in the report, the significance level at which they are conducted should be stated explicitly.
- Sampling errors for the principal estimates in a report should always be available to the reader; thus, tables of sampling errors, design effects, or generalized variance functions should be readily accessible, either through a presentation in the printed report or electronically on the Internet.
- When space limitations preclude publishing detailed information, relevant technical publications should be provided as references, both print and electronic (URL) references.

Since details concerning the development of sampling errors are often of interest to more technically oriented users, these details can be left to technical reports or user's manuals. Consequently, the subcommittee recommends the following for inclusion in technical reports or user's manuals:

- The method used for calculating sampling error should be identified, with reference to a more detailed description. If generalized models are used to provide sampling errors, the data user should have access to the models, the assumptions underlying the models, and references to the results of the modeling.
- Sampling error calculations for different types of estimates should be described (e.g., levels, percents, ratios, means, and medians).
- Evaluations of the procedures used to estimate sampling errors should be described and discussed.

4.4.2. Reporting error sources in analytic reports: nonresponse error

Nonresponse error is an error of nonobservation, reflecting an unsuccessful attempt to obtain the desired information from an eligible unit. Nonresponse reduces sample size, results in increased variance, and introduces a potential for bias in the survey estimates. Nonresponse rates are frequently reported and are often viewed as a proxy for the quality of the survey. Nonresponse rates may be calculated differently for different purposes (Lessler and Kalsbeek 1992; Gonzalez, Kasprzyk, and Scheuren 1994; Council of American Survey Research Organizations 1982; American Association for Public Opinion Research 2000) and they are often miscalculated. The complexities of the survey design often make calculation and communication of response rates confusing and potentially problematic. While reporting nonresponse rates is important, nonresponse rates alone provide no indication of nonresponse bias. Special studies are necessary.

In the course of reviewing errors due to nonresponse, a number of areas were identified that need to be addressed when reporting results in an agency analytic report:

- Unit and item nonresponse should be identified as important sources of error.
- Overall unit response rates (weighted using base weights, and unweighted) should be provided, as well as definitions of the response rates given; this includes providing definitions of the numerator and denominator in the response rate calculation.
- In multistage designs, weighted (using base weights) and unweighted response rates at each interview level should be given, and an overall response rate computed. Assumptions necessary for the response rate calculation should be stated.
- Longitudinal surveys should report separately the response rate for the first wave of the survey and each follow-up wave of the survey, as well as the cumulative response rate. Reporting other response rates is encouraged; for example, reporting the response rate of sample units who responded in all waves of a longitudinal survey may be informative to the longitudinal data analyst.
- Subgroup response rates should be provided if specific subgroups are important to the analysis.
- Item response rates should be summarized and items with low response rates identified.
- Unit and item nonresponse adjustment procedures – whether they are weighting procedures or imputation methods – should be identified.
- When unit or item nonresponse rates are lower than the agency deems ''reasonable,'' special studies to assess the bias due to nonresponse should be conducted, the results summarized, and the detailed report referenced.
- If available, studies designed to measure potential nonresponse bias should be referenced.

There is typically more information about errors of nonresponse than there is about other error sources. Consequently, technical reports and user's manuals are the most reasonable types of reports to document the myriad important details related to nonresponse as a source of error in surveys. Details that ought to be reported in a technical report or user's manual include the following:

- Procedures used to compensate for missing data, both unit and item nonresponse should be described and the key variables used in the procedures identified. The effects of these procedures (if known) on the estimates should be discussed.

- Evaluations of the missing data procedures, both unit and item, should be conducted, the results summarized, and a detailed report referenced.
- Special studies that aim to understand or measure the bias due to nonresponse should be conducted and then summarized, and the detailed report referenced.
- Steps taken to maximize the response rate and the extent of nonresponse follow-up should be described.
- Subgroup response rates for key subpopulations should be calculated and made available.
- Reasons for nonresponse (refusals, noncontacts, etc.) should be monitored and reported.

4.4.3. Reporting error sources in analytic reports: coverage error

Coverage error includes both undercoverage and overcoverage. It is the error in the estimate that results from the failure to include some population units on the frame used for sample selection (undercoverage) and the error associated with the failure to identify units represented on the frame more than once or units that are not members of the target population (overcoverage). The source of coverage error is the sampling frame itself. It is important, therefore, to have information about the quality of the sampling frame and its completeness for the target population. Measurement methods for coverage error often rely on methods external to the survey operations; for example, comparing survey estimates to independent sources, or implementing a case-by-case matching of two lists. Other methods for measuring coverage error may use information from the sample; that is, coverage error can be identified using sample statistics, such as the birth rate (new unit rate), death rates (out-of-business rate in establishment surveys), duplication and misclassification rates and comparing whether these rates are significantly higher or lower than comparative data (U.S. Federal Committee on Statistical Methodology 1988).

The nature of the publication and survey (one-time versus continuing) plays a significant role in determining what and how much an analyst reports about this source of error. For analytic reports, the following areas and topics related to coverage error should be reported:

- Coverage error should be mentioned explicitly as a source of nonsampling error.
- The target population, the set of elements about which information is wanted and parameter estimates required, and the frame population, the set of elements either listed directly as units in the frame or identified through a more complex frame concept, should be defined and a clear statement made about exclusions in the frame population. The estimated percent of the excluded frame population should be provided.
- The sampling frame should be identified and described. Information about the frame should be reported, such as the year the frame was developed, whether the frame has changed over time, whether it has been updated for births, deaths, and other relevant changes to the target population, and whether gaps or other problems in the frame exist that would affect its quality.
- If available, an overall coverage rate should be defined and supplied to the user.
- References to studies about the sampling frame, its quality, and issues related to coverage should be included.
- If known, the effect of coverage error on key survey estimates should be reported.

Detailed reporting on the methods and procedures related to measuring coverage error does not usually occur in technical appendices of analytic reports. Because of their essentially unrestricted length, technical reports and user's manuals are better venues for reporting additional information about coverage error. Topics that ought to be addressed in these publications include:

- A general assessment of the quality of the sampling frame should be supplied to the data user. This should include a description and discussion of the limitations of the frame.
- Procedures used to update the frame should be described.
- Subpopulation coverage rates should be reported, particularly if the subpopulations are important analytic domains.
- Poststratification procedures and the effects of using such procedures should be described.
- A summary of results from studies that aim to measure coverage error should be provided.

4.4.4. Reporting error sources in analytic reports: measurement error

Measurement error is characterized as the difference between the observed value of a variable and the true, but unobserved, value of that variable. Measurement error comes from four primary sources in survey data collection: the interviewer, as the deliverer of the questions; the questionnaire, as the official presentation or request for information; the data collection method, as the way in which the request for information is made; and the respondent, who receives the request. These sources comprise the entirety of data collection, and each source can introduce error into the measurement process. For example, measurement error may occur in respondents' answers to survey questions, including misunderstanding the meaning of the question, failing to recall the information accurately, and failing to construct the response correctly (e.g., by summing the components of an amount incorrectly). Measurement errors are difficult to quantify, usually requiring special, expensive studies, such as reinterview programs, record check studies, behavior coding, cognitive testing, and randomized experiments.

Because of the difficulty, complexity, and expense of quantifying measurement error, publications typically do not report very much detail on this particular error source. However, a data user cannot understand the limitations of the data – from a measurement error point of view – unless the data collection program takes steps to explicitly provide such information. The studies required are costly, time-consuming, and not available quickly. Recommendations for reporting this error source must take the practical realities into consideration. Thus, for analytic reports, the following is recommended:

- Measurement error, in general, ought to be defined and described as a source of nonsampling error.
- Authors should provide examples of different sources of measurement error likely to be found in the survey.
- The report should include brief summaries (if available) of studies to quantify and understand measurement error in the context of the survey, such as reinterviews,

record check studies, or split-sample experiments. Reference should be made to more detailed methodological or technical reports.

- The amount of information and detail reported on sources of measurement error is related to the relative importance of the source of error, what is known about the source of error, and how it may affect characteristics analyzed in the report.
- Indirect data quality indicators should be reported, such as steps taken to reduce measurement error (for example, pretests, experiments, interviewer training, and cognitive testing of questionnaires).
- Reports should refer to synthesis reports or quality profiles that describe the variety of measurement studies conducted, their results, and the possible effects on analysis.
- In general, the amount of detail reported concerning sources of measurement error should depend on the known or assumed effects of the error source on key statistics.

Technical reports, user's manuals, and quality profiles are the appropriate dissemination venues for detailed reporting on the planning, conduct, and analysis of measurement error studies. These reports are most useful to the data user if the results can be related to the key statistics and findings of the analytic report.

4.4.5. Reporting error sources in analytic reports: processing error

Processing error occurs after the survey data are collected, during the processes that convert reported data to published estimates and consistent machine-readable information. Processing errors are of two types: errors *identified* during the data processing phase and errors *made* during the processing phase. Each processing step, from data collection to the publication of the final survey results, can identify or generate errors in the data or in the published statistics. These errors range from a simple recording error, that is, a transcribing or transmission error, to more complex errors arising from a poorly specified edit or imputation model. They tend not to be well reported or well documented, and are seldom treated in the survey research literature. Processing errors include data entry, coding, and editing and imputation errors. In addition, errors in computer programs can also be considered as a source of processing error (U.S. Federal Committee on Statistical Methodology 1988). Programming errors are detected and controlled by using test files, that is, simulating problems that may occur in actual data files, and by a review of the program code. Imputation errors are included under processing error because many agencies treat failed edits as missing, and impute values for them.

Error rates are determined through quality control samples; however, in recent years authors have advocated continuous quality management practices (Morganstein and Marker 1997; Linacre and Trewin 1989). Performance statistics, such as data keying error rates, edit failure rates, imputation rates, and coding error rates are usually produced for management purposes during the data processing operations. These statistics, collected during the data processing stages of the survey cycle, however, provide no measure of the effect of the errors on the accuracy of the data. Rather they provide information to survey managers about the data collection and post survey operations where survey operations can be improved through improved measurement (a high edit failure rate may suggest the need for improved questionnaires and question wording, for example) or improved training (to reduce coding errors, for example).

Analytic reports provide limited information about processing error to the general public. Considerable detail is required to describe the data processing operations and certain performance statistics. While this information is important to survey practitioners and methodologists, it is usually considered to be too detailed for inclusion in an agency analytic report. The performance statistics taken together can provide a broad understanding of the quality of the survey operations. However, since many of the details are felt to be beyond the scope of interest of the primary target audience of an analytic report, these statistics and the description of processing operations should be published in a separate volume and referenced by the analytic report. The following is recommended regarding an analytic report:

- Errors in data processing ought to be described as a potential source of error in surveys.
- Data keying error rates, scanning error rates, other data entry error rates, coding error rates, and edit failure rates should be referenced as available on the Internet or in technical reports, user's manuals, or quality profiles – particularly for key variables.
- A short discussion of the quality control aspects of the data processing operations should be provided to the user.
- Processing error studies, such as coder-variance studies, should be referenced in the report and be available to the user community electronically or through technical reports, user's manuals, or quality profiles.

As discussed above, appendices to analytic reports are not the most appropriate venue for providing detailed technical information. Technical and methodological reports provide more opportunities to disseminate detailed information on all aspects of data processing operations. Consequently, these reports ought to provide the following with regard to survey data processing:

- Description of the quality control aspects of survey data processing and the data input operations, such as data keying and imaging.
- Quality control results related to data entry, coding, and editing.
- Description of processing error studies, particularly coder-variance studies, with the results summarized and implications for analysis (if any) clarified.
- Discussion of the extent to which data processing operations alter responses, particularly in the edit and imputation phases. The altered data should be identified for the end-user.
- Identification and description of changes in processing operations in continuing and periodic surveys.

4.4.6. Reporting error sources in analytic reports: total survey error

Total Survey Error describes a combination of all the above forms of error. Total survey error is typically reported as a compilation of reports about the impact on survey data of sampling error and each of the different types of nonsampling error. To date, very few data collection programs have successfully integrated different sources of error into a single computation to determine the effect of the errors on key statistics.

The study of total survey error is important in that it can improve the conduct and analysis of federal government surveys. Understanding the individual sources of error

in surveys and their contribution to the overall error is key to improving the design and implementation of surveys. With some understanding of the combined effects of error sources on the data, researchers are better able to choose analytic methods that are most robust against such errors. Because of the practical difficulty of reporting total survey error, a number of good practices now reflected in the practices of many federal statistical agencies, should be reemphasized:

- Continuing and periodic survey programs ought to regularly report summary results obtained from methodological studies; implications of these results for analysis should be addressed. One way to address this is through the regular production of quality profiles for continuing and periodic survey programs.
- Lacking a report on all error sources, survey programs should identify the prominent sources of error, report on them individually in technical reports, and discuss the implications for analysis.
- Survey programs ought to define, routinely calculate, monitor, discuss, and report survey performance indicators to data users.
- Survey programs should allocate a portion of their budgets to designing, implementing, and reporting on methodological studies that help users understand sources of error in the survey and implications for analysis.
- Technical or methodology reports describing the sample design, estimation, and data collection procedures should be available for data users. Such reports allow users to judge the quality of the survey and its operations.
- Survey programs should compare aggregate results to other comparable data, assess reasons for differences, and report on the results of comparisons.
- Survey programs should actively develop Web applications that allow the user to easily obtain information concerning the overall quality of the data. Continuing and periodic surveys should develop long-range research plans that systematically address the measurement of the components of total survey error.

The classification of error sources in surveys, as described above, provides a framework for users of statistical data to develop an understanding of the nature of the data they analyze. An understanding of the limitations of data can assist an analyst in developing methods to compensate for any known shortcomings of the data.

5. Reporting Error Sources on the Internet

The Internet has become the principal medium for the dissemination of data products for most U.S. federal statistical agencies. Because of its importance as a dissemination tool and its unique characteristics and capabilities, the subcommittee examined it as a separate report format.

5.1. Reporting error sources on the Internet: results of a study

A third Federal Committee on Statistical Methodology study (Giesbrecht, Miller, Moriarity, and Ware-Martin 1999) reviewed guidelines and practices for reporting error sources over the Internet. Some federal agencies have written standards for Web sites, but these generally focus on Web site design, layout, and administrative access. A few

agencies, such as the U.S. Census Bureau, have begun the process of developing standards for providing information about data quality over the Internet (U.S. Bureau of the Census 1997). This draft report gives details of data quality information that ought to be provided for the user, but does not require or suggest the use of Internet features for making information more accessible. Generally, standards documents related to Internet practices reiterate standards for printed documents (for example, United Nations Economic and Social Council 1998).

The study reviewed the accessibility of data quality documentation on current Internet sites of 14 federal agencies with survey data collections. Online data documentation was available for most of the sites visited (78 percent). For about one-half the sites, offline documentation was referenced as well. Most agencies seem to upload their printed reports and documentation in the form of simple text or Adobe Acrobat format files. In addition, one-half of the sites offered technical support online and an additional 29 percent included lists of telephone contacts on their Web sites. The study also noted a few best practices found on the visited Web sites, such as the availability of pop-up windows providing definitions of column and row headings in tables, links to send e-mail messages to technical specialists, links to ''survey methodology'' and ''survey design'' documentation, explicit directions to users about errors and comparability issues, links from one agency's home page to another's, and common access points for statistical information.

The study found that current Internet standards for data quality information echo the standards for printed reports and statistical tables. More explicit guidelines for how the advantages of the Internet media should be employed to make data quality information more accessible do not seem to exist. However, the development of metadata standards (Dippo 1997) as an integral part of the survey measurement process may facilitate the creative use of the Internet.

5.2. Reporting error sources on the Internet: discussion and recommendations

The use of the Internet for reporting statistical information is growing and evolving so fast that recommendations seem inappropriate since they become out-of-date very quickly. The potential use of this new medium has not been fully developed, and statistical agencies, while providing much information on the Internet, have only begun to explore its potential. In general, agencies report electronically what is reported on paper, often in the form of Adobe Acrobat format files that are no more interactive than a paper report. Thus, the limitations on reporting information on the data collection program and sources of error are limitations of the printed report itself.

Large gaps exist between the potential of the medium and implementation within the medium. The key issue is how to organize and display statistical information and its corresponding documentation in a way that can be understood and easily accessed by the user community. Thus, it is important for statistical agencies to continue developing online design features, such as frames, audio/video, hyperlinks to relevant documents (such as design, estimation, and technical documentation) or parts of the same document, pop-up windows (for, among other applications, providing data concepts and definitions or the sampling error of an estimate in a table), on-line data analysis tools, user forums, and e-mail technical support links to improve service to data users.

The Subcommittee recommends that:

> *Agencies should systematically and regularly review, improve access to, and update reports and data products available on the Internet, particularly reports about the quality of the data; the amount of information about data quality should be no less than that contained in printed reports; linkage features available on the Internet, such as hypertext links, should be used to improve access to information about the data collection program and its sources of error. Information displayed on the Internet should incorporate good design principles and ''best practices'' of displaying data and graphics on the Web.*

Agency practices will dictate whether the Internet reporting function is decentralized or not. Either way, financial and staff resources should be allocated to developing new applications to improve online access to information about the quality of data in reports and products on the Internet.

As printing and traditional dissemination costs continue to increase and Internet access grows, it will become increasingly common to find information available *only* through the Internet. Internet dissemination ought to spur the development of new ways to present statistical information, and new ways to inform data users about the quality of this statistical information. At this time, however, based on the review of Internet sites, the paper report model is almost universal, resulting in an Internet product developed only after the paper product is completed. The potential of the Internet to present and display information has not been addressed solely with regard to the features and capabilities of the Internet – otherwise, the use of video, audio, frames, hyperlinks, and other features of the interface would be more obvious. Data, reports, and press releases available only through the Internet should be developed to take maximum advantage of the new medium.

6. Conclusion

U.S. statistical agencies need to better address the difficult task of informing data users about the myriad dimensions of data quality. Many U.S. statistical agencies have written policies for informing data users about various aspects of the quality of data. These policies follow from a professional ethic of openness about the process of data collection and analysis, without which data used to inform public policy decisions would not be credible. However, as we have seen, data quality information that users need is not always made available and is not consistently reported across agencies. While the diversity of publications and the decentralized structure of the U.S. statistical system may make it difficult to institute consistent and comprehensive reporting of data quality content, it is important for agencies to maintain a dialogue with each other to improve reporting practices, to pay attention on a continuous basis to the data users' need for information, and place high priority on the provision of such information through the various dissemination media.

7. References

American Association for Public Opinion Research (2000). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. Ann Arbor, Michigan: American Association for Public Opinion Research.

Atkinson, D., Schwanz, D., and Sieber, W.K. (1999). Reporting Sources of Error in

Analytic Publications. Seminar on Interagency Coordination and Cooperation. U.S. Office of Management and Budget (Statistical Policy Working Paper 28*),* 329–341. Available at http://www.fcsm.gov.

Brackstone, G. (1999). Managing Data Quality in a Statistical Agency. Survey Methodology. 25, 139–149.

Carson, C. and Liuksila, C. (2001). Further Steps Toward a Framework for Assessing Data Quality. Paper presented at the International Conference on Quality in Official Statistics. Available on CD-Rom. Stockholm: Statistics Sweden.

Colledge, M. and March, M. (1997). Quality Policies, Standards, Guidelines, and Recommended Practices at National Statistical Agencies. In Survey Measurement and Process Quality, L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz, and D. Trewin (eds). New York: John Wiley, 501–522.

Collins, M. and Sykes, W. (1999). Extending the Definition of Survey Quality. Journal of Official Statistics, 15, 57–66.

Council of American Survey Research Organizations (1982). On the Definitions of Response Rates. Port Jefferson, NY.

Davies, P. and Smith, P. (eds) (1998). Model Quality Report in Business Statistics (four volumes). Eurostat. Final Report of SUPCOM Project 1997, lot 6.

Dippo, C. (1997). Survey Measurement and Process Improvement. In Survey Measurement and Process Quality: Concepts and Integration, L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz, and D. Trewin (eds). New York: John Wiley, 457–474.

Eurostat (2000a). Definition of Quality in Statistics. Eurostat Working Group on Assessment of Quality in Statistics. Luxembourg, 4–5 April.

Eurostat (2000b). Standard Quality Report. Eurostat Working Group on Assessment of Quality in Statistics. Luxembourg, 4–5 April.

Giesbrecht, L., Miller, R., Moriarity, C., and Ware-Martin, A. (1999). Reporting Data Quality on the Internet. Seminar on Interagency Coordination and Cooperation, U.S. Office of Management and Budget (Statistical Policy Working Paper 30), 342–354. Available at http://www.fcsm.gov.

Gonzalez, M.E., Kasprzyk, D., and Scheuren, F. (1994). Nonresponse in Federal Surveys: An Exploratory Study. Amstat News, 208, Alexandria, VA: American Statistical Association.

Gonzalez, M.E., Ogus, J.L., Shapiro, G., and Tepping, B.J. (1975). Standards for Discussion and Presentation of Errors in Survey and Census Data. Journal of the American Statistical Association, 70, 5–23.

Kasprzyk, D., McMillen, M., Atkinson, D., Giesbrecht, L., Schwanz, D., and Sieber, W.K. (1999). Reporting Sources of Error: The United States Experience. Proceedings of the 52nd Session of the International Statistical Institute. International Association of Survey Statisticians, 19–30.

Kish, L. (1965). Survey Sampling. New York: John Wiley.

Lessler, J.T. and Kalsbeek, W.D. (1992). Nonsampling Error in Surveys. New York: John Wiley.

Linacre, S. and Trewin, D. (1989). Evaluation of Errors and Appropriate Resource Allocation in Economic Collections. Proceedings of the Annual Research Conference. U.S. Bureau of the Census, Washington, DC, 197–209.

Lyberg, L. et al. (2001). Summary Report from the Leadership Group (LEG) on Quality. Eurostat and Statistics Sweden, Stockholm, 31 July.

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

McMillen, M. and Brady, S. (1999). Reporting Sources of Error in Short Format Publications. Seminar on Interagency Coordination and Cooperation. U.S. Office of Management and Budget (Statistical Policy Working Paper 28), 316–328. Available at http://www.fcsm.gov.

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

Statistics Canada (2002). Symposium 2001: Achieving Data Quality in a Statistical Agency: A Methodological Perspective. Ottawa, Canada.

Statistics Canada (1991). Symposium 90: Measurement and Improvement of Data Quality. Ottawa, Canada.

Statistics Sweden (2001a). International Conference on Quality in Official Statistics. Stockholm. Available on CD-rom.

Statistics Sweden (2001b). Quality Definition and Recommendations for Quality Declarations of Official Statistics. Reports on Statistical Coordination for the Official Statistics of Sweden, MIS 2001 No. 1. Stockholm.

United Nations Economic and Social Council (1998). Guidelines for Statistical Metadata on the Internet. Paper contributed to the Conference of European Statisticians, Forty-Sixth Plenary Session, Paris, May 18–20.

U.S. Bureau of the Census (1997). Survey Design and Statistical Methodology Metadata IT Standards (draft). Washington, DC: U.S. Department of Commerce.

U.S. Federal Committee on Statistical Methodology (2001a). Measuring and Reporting Sources of Error in Surveys. Washington, DC: U.S. Office of Management and Budget. (Statistical Policy Working Paper 31). Available at http://www.fcsm.gov.

U.S. Federal Committee on Statistical Methodology (2001b). 2001 Federal Committee on Statistical Methodology Research Conference. Washington, DC: U.S. Office of Management and Budget. (Statistical Policy Working Paper 34). Available at http://www.fcsm.gov.

U.S. Federal Committee on Statistical Methodology (2000). 1999 Federal Committee on Statistical Methodology Conference: Complete Proceedings. Washington, DC: U.S. Office of Management and Budget (Statistical Policy Working Paper 30). Available at http://www.fcsm.gov.

U.S. Federal Committee on Statistical Methodology (1999). Seminar on Interagency Coordination and Cooperation. Washington, DC: U.S. Office of Management and Budget (Statistical Policy Working Paper 28). Available at http://www.fcsm.gov.

U.S. Federal Committee on Statistical Methodology (1997). Seminar on Statistical Methodology in the Public Service. Washington, DC: U.S. Office of Management and Budget (Statistical Policy Working Paper 26). Available at http://www.fcsm.gov.

U.S. Federal Committee on Statistical Methodology (1991). Seminar on Quality of Federal

Data. Washington, DC: U.S. Office of Management and Budget (Statistical Policy Working Paper 20). Available at http://www.fcsm.gov.

U.S. Federal Committee on Statistical Methodology (1988). Quality in Establishment Surveys. Washington, DC: U.S. Office of Management and Budget (Statistical Policy Working Paper 15). Available at http://www.fcsm.gov.

Received November 2001
Revised August 2003