Chapter XI
Reporting and compensating for non-sampling errors for surveys in Brazil: current practice and future challenges

Pedro Luis do Nascimento Silva
Escola Nacional de Ciências Estadísticas/
Instituto Brasileiro de Geografia e Estatística
(ENCE/IBGE)
Rio de Janeiro, Brazil

Abstract

The present chapter discusses some current practices for reporting and compensating for non-sampling errors in Brazil, considering three classes of errors: coverage errors, non-response, and measurement and processing errors. It also identifies some factors that make it difficult to focus greater attention on the measurement and control of non-sampling errors. In addition, it identifies some recent initiatives that might help to improve the situation.

Key terms: survey process, coverage, non-response, measurement errors, survey reporting, data quality.

A. Introduction

1. The notion of error as applied to a statistic or estimate of some unknown target quantity (or parameter) must be defined. It refers to the difference between the estimate (say, $\hat{Y}$) and the theoretical "true parameter value" (say, Y) that would be obtained or reported if all sources of error were eliminated. Perhaps, as argued by some, a better term would be deviation (see discussion in Platek and Särndal (2001, sect. 5). However, the term error is so entrenched that we shall not attempt to avoid it. Here, we are concerned with survey errors, that is to say, errors of estimates based on survey data. According to Lyberg and others (1997, p. xiii), "survey errors can be decomposed in two broad categories: sampling and non-sampling errors". The discussion of survey errors, in modern terminology, is part of the wider discussion of data quality.

2. To illustrate the concept, suppose that the estimate of the average monthly income for a certain population reported in a survey is 900 United States dollars, and that the actual average monthly income for members of this population, obtained from a complete enumeration without errors of reporting and processing, is US$ 850. Then, in this example, the error of the estimate would be US$ +50. In general, survey errors are unobserved, because the true parameter values are unobserved (or unobservable). One instance in which at least the sampling errors of statistical estimates may be observed is that provided by sampling from computer records, where the differences between estimates and the values computed using the full data sets can then be computed, if required. Public use samples of records from a population census provide an example of practical application. In Brazil, samples of this type have been selected from population census records since 1970. However, situations like this are the exception, not the rule.

3. Sampling errors refer to differences between estimates based on a sample survey and the corresponding population values that would be obtained if a census was carried out using the same methods of measurement, and are "caused by observing a sample instead of the whole population" (Särndal, Swensson and Wretman, 1992, p. 16). "Non-sampling errors include all other errors" (ibid.) affecting a survey. Non-sampling errors can and do occur in all sorts of surveys, including censuses. In censuses and in surveys employing large samples, non-sampling errors are the main source of error that one must be concerned with.

4. Survey estimates may be subject to two types of errors: bias and variable errors. Bias refers to errors that affect the expected value of the survey estimate, taking it away from the true value of the target parameter. Variable errors affect the spread of the distribution of the survey estimates over potential repetitions of the survey process. Regarding sampling errors, bias is usually avoided or made negligible by using adequate sampling procedures, sample size and estimation methods. Hence, the spread is the main aspect of the distribution of the sampling error that one has to consider. A key parameter describing this spread is the standard error, namely, the standard deviation of the sampling error distribution.

5. Non-sampling errors include two broad classes of errors (Särndal, Swensson and Wretman, 1992, p. 16): “errors due to non-observation” and “errors in observations”. Errors due to non-observation result from failure to obtain the required data from parts of the target population (coverage errors) or from part of the selected sample (non-response error). Coverage or frame errors refer to wrongful inclusions, omissions and duplications of survey units in the survey frame, leading to over- or undercoverage of the target population. Non-response errors are those caused by failure to obtain data for units selected for the survey. Errors in observations can be of three types: specification errors, measurement errors and processing errors. Biemer and Fecso (1995, chap. 15) define specification errors as those that occur when “(1) survey concepts are unmeasurable or ill-defined; (2) survey objectives are inadequately specified; or (3) the collected data do not correspond to the specified concepts or target variables”. Measurement errors concern having observed values for survey questions and variables after data collection that differ from the corresponding true values that would be obtained if ideal or gold standard measurement methods were used. Processing errors are those introduced during the processing of the collected data, that is to say, during coding, keying, editing, weighting and tabulating the survey data. All of these types of errors are dealt with in the subsections of section B, with the exception of specification errors. The exclusion of specification errors from our discussion does not mean that they are not important, but only that discussion and treatment of these errors are not well established in Brazil.

6. Other approaches to classifying non-sampling errors are discussed in a United Nations manual (see, United Nations, 1982). In some cases, there is no clear dividing line between non-response, coverage and measurement errors, as is the case in a multistage household sample survey when a household member is missed in an enumerated household: Is this a measurement error, a non-response or a coverage problem?

7. Non-sampling errors can also be partitioned into non-sampling variance and non-sampling bias. Non-sampling variance measures the variation in survey estimates if the same sample would be submitted to hypothetical repetitions of the survey process under the same essential conditions (United Nations, 1982, p. 20). Non-sampling bias refers to errors that result from the survey process and survey conditions, and would lead to survey estimates with an expected value different from the true parameter value. As an example of non-sampling bias, suppose that individuals in a population tend to underreport their income by an average 30 per cent. Then, irrespective of the sampling design and estimation procedures, without any external information, the survey estimates of average income would be on average 30 per cent smaller than the true value of the average income for members of the population. Most of the discussion in the present chapter deals with avoiding or compensating for non-sampling bias.

8. Data quality issues in sample surveys have received increased attention in recent years, with a number of initiatives and publications addressing the topic, including several international conferences (see sect. D). Unfortunately, the discussion is still predominantly restricted to developed countries, with little participation and contribution coming from developing and transition countries. This is the main conclusion one reaches after examining the proceedings and publications issued after these various conferences and initiatives. However, several papers have recently been published on this topic in respect of surveys in transition countries in the journal

Statistics in Transition (Kordos, 2002), but this journal does not appear to have wide circulation in libraries across the developing world.

9. Regarding sampling errors, a unified theory of measurement and estimation exists [see, for example, Särndal, Swensson and Wretman (1992)], which is supported by the widespread dissemination of probability sampling methods and techniques as the standard for sampling in survey practice (Kalton, 2002), and also by standard generalized software that enables practical application of this theory to real surveys. If samples are properly taken and collected, estimates of the sampling variability of survey estimates are relatively easy to compute. This is already being done for many surveys in developing and transition countries, although this practice is still far from becoming a mandatory standard.

10. The dissemination and analysis of such variability measures lag behind, however. In many surveys, sampling error estimates are neither computed nor published, or are computed/published only for a small selection of variables/estimates. Generally, they are not available for the majority of the survey's estimates because such a massive computational undertaking is involved. While this may make it difficult for an external user to assess the degree of sampling variability for a particular variable of interest, it is possible nevertheless to gauge its order of magnitude by comparing it with a similar variable for which the standard error was estimated. Commentary about survey estimates often ignores the degree of variability of the estimates. For example, the Brazilian Monthly Labour Force Survey (Instituto Brasileiro de Geografia e Estatística, 2002b), started in 1980, computes and publishes every month estimates of the coefficients of variation (CVs) of the leading indicators estimated from the survey. However, no estimates of standard errors are computed for differences of such indicators between successive months, or months a year apart. Yet, most of the survey commentary published every month together with the estimates is about change (variations in the monthly indicators). Only very recently were such estimates of standard errors for estimates of change computed for internal analysis [see Correa, Silva and Freitas (2002)], and these are not yet made available regularly for external users of survey results. The same is true when the estimates are "complex", as is the case with seasonally adjusted series of labour-market indicators.

11. If the situation is far from ideal regarding sampling errors, where both theory and software are widely available, and a widespread dissemination of the sampling culture has taken place, treatment of non-sampling errors in household and other surveys in developing countries is much less developed. Lack of a widely accepted unifying theory [see Lyberg and others (1997, p. xiii); Platek and Särndal (2001)]; and subsequent discussion), lack of standard methods for compiling information about and estimating parameters of the non-sampling error components, and lack of a culture that recognizes the importance of measuring, assessing and reporting on these errors imply that non-sampling errors, and their measurement and assessment, receive less attention in surveys carried out in developing or transition countries. This is not to say that most surveys carried out in developing or transition countries are of low quality, but rather to stress that we know little about their quality levels.

12. With this background information on the status of the non-sampling error measurement and control for surveys carried out in developing and transition countries, we move on to discuss the status of current practice (sect. B) regarding the Brazilian experience. Although limited to

what is found in one country (Brazil), we believe that this discussion is relevant for statisticians in other developing countries, given that literature on the subject is scarce. We then indicate what challenges lie ahead for improved survey practice in developing and transition countries (sect. C), again from the perspective of survey practice in Brazil.

B. Current practice for reporting and compensating for non-sampling errors in household surveys in Brazil

13. In Brazil, the main regular household sample surveys with broad coverage are carried out by Instituto Brasileiro de Geografia e Estatística (IBGE), the Brazilian central statistical institute. To help the reader understand the references to these surveys, we present their main characteristics, coverage and periods in table XI.1.

Table XI.1. Some characteristics of the main Brazilian household sample surveys

Survey name	Period	Population coverage	Topic/theme
Population Census	Every 10 years (latest in 2000)	Residents in private and collective households in the country	Household items, marital status, fertility, mortality, religion, race, education, labour, income
National Household Sample Survey (PNAD)	Annual, except for census years	Residents in private and collective households in the country, except in rural areas of northern region	Household items, religion, race, education, labour, income and special supplements on varied topics
Monthly Labour Force Survey (PME)	Monthly	Residents in private households in six large metropolitan areas	Education, labour, income
Household Expenditure Survey (POF)	1974-1975, 1986-1987, 1995-1996, 2002-2003	National in the 2002-2003 edition; 11 large metropolitan areas in two previous editions; national in 1974-1975 edition	Household items, family expenditure and income
Living Standards Measurement Survey (PPV)	1996-1997	Residents in private households in the north-east and south-east regions	Extensive coverage of topics relating to measurement of living standards
Urban Informal Economy Survey (ECINF)	1997	Residents involved in the informal economy in private households in urban areas	Labour, income and characteristics of business in the informal economy

1. Coverage errors

14. Coverage errors refer to under- or overcoverage of survey population units. Undercoverage occurs when units in the target population are omitted from the frame, and thus would not be accessible for the survey. Overcoverage occurs when units not belonging to the target population are included in the frame and there is no way to separate them from eligible units prior to sampling, as well as when the frame includes duplicates of eligible units. Coverage errors may also refer to wrongful classification of survey units in strata due to inaccurate or outdated frame information (for example, when a household is excluded from the sampling process for not being occupied, when in fact it was occupied at the time the survey was carried out). Undercoverage is usually more damaging than overcoverage with respect to the estimates from a survey. There is no way we can recover missing units but units outside the universe can often be identified during the fieldwork or data processing and appropriately corrected or adjusted; the units outside the universe do, however, result in increased survey cost per eligible unit.

15. Coverage problems are often considered more important when a census is carried out than when a sample survey is carried out because, in a census, there are no sampling errors to worry about. However, this is a misconception. In some sample surveys, coverage can sometimes be as big a problem as sampling error, if not bigger. For example, sample surveys can sometimes exclude from the sampling process (hence giving them zero inclusion probability) units in certain hard-to-reach areas or in categories that are hard to canvass. This may occur for reasons of interviewer safety (for example, where surveying would involve areas of conflict or high-level violence) or of cost (for example, when travelling to parts of the territory for interviewing is prohibitively expensive or takes too long). If the definition of the target population does not describe such exclusions precisely, the resulting survey will lead to undercoverage problems. Such problems are likely to affect estimates in terms of bias, since the units excluded from the survey population will tend to be different from those that are included. When the survey intends to cover such hard-to-reach populations, special planning is required to make sure that the coverage is extended to include these groups in the target population, or the population for which inferences are to be drawn.

16. A related problem arises with some repeated surveys carried out in countries with poor telephone coverage and perhaps high illiteracy rates, where data collection must rely on face-to-face interviews. When these surveys have a short interviewing period, their coverage may often be restricted to easy-to-reach areas. In Brazil, for example, the Monthly Labour Force Survey (PME) is carried out in only six metropolitan areas (Instituto Brasileiro de Geografia e Estatística, 2002b). Its limited definition of the target population is one of the key sources of criticism of the relevance of this survey: with a target population that is too restricted for many uses, it does not provide information on the evolution of employment and unemployment elsewhere in the country. Although the survey correctly reports its figures as relating to the "survey population" living in the six metropolitan areas, many users wrongly interpret the figures for the sum of these six areas as if they relate to the overall population of Brazil. Redesign of the survey is planned in order to address this issue in 2003-2004. Similar issues arise in other surveys like, for example, the Brazilian Income and Expenditure surveys of 1987-1988 and 1995-1996 (coverage restricted to 11 metropolitan areas) and the Brazilian Living Standards

Measurement Study (LSMS) survey of 1996-1997 (coverage restricted to the north-east and south-east regions only). To a lesser degree, this is also the case with the major "national" annual household sample survey carried out in Brazil (Instituto Brasileiro de Geografia e Estatística, 2002a). This survey does not cover the rural areas in the northern region of Brazil owing to prohibitive access costs. Bianchini and Albieri (1998) provide a more detailed discussion of the methodology and coverage of various household surveys carried out in Brazil.

17. Similar problems are experienced by many surveys in other developing and transition countries, where the coverage of some hard-to-reach areas of the country on a frequent basis may be too costly. An important rule to follow regarding this issue is that any publication based on a survey should include a clear statement about the population effectively covered by that survey, followed by a description of potentially relevant subgroups that have been excluded from it, if applicable.

18. Coverage error measures are not regularly published together with survey estimates to allow external users an independent assessment of the impact of coverage problems in their analyses. These measures may be available only when population census figures are published every 10 years or so and, even in this case, they are not directly linked to the coverage problem of the household surveys carried out in the preceding decade.

19. In Brazil, the only "survey" where more comprehensive coverage analysis is carried out is the population census. This is usually accomplished by a combination of post-enumeration sample surveys and demographic analysis. A post-enumeration sample survey (PES) is a survey carried out primarily to assess coverage of a census or similar survey, though in many country applications, the PES is often used to evaluate survey content as well. In Brazil, the PES following the 2000 population census sampled about 1,000 enumeration areas and canvassed them using a separate and independent team of enumerators who had to follow the same procedures as those followed by the regular census enumerators. After the PES data are collected, matching is carried out to locate the corresponding units in the regular census data. Results of this matching exercise are then used to apply the dual-system estimation method [see, for example, Marks (1973)], which produces estimates of undercoverage such as those reported in table XI.2 below. Demographic analysis of population stocks and flows based on administrative records of births and deaths can also be used to check on census population counts and assess their degree of coverage. In Brazil, this practice is fruitful only in some States in the south and south-east regions, where records of births and deaths are sufficiently accurate to provide useful information for this purpose.

20. A serious impediment towards generalized application of PES surveys for census coverage estimation and analysis is their high cost. These surveys need to be carefully planned and executed if their results are to be reliable. Also, it is important that they provide results disaggregated to some extent, or otherwise their usefulness will be quite limited. In some cases, the resources that would be needed for such a survey are not available, and in others, census planners may believe that those resources would be better spent in improving the census operation itself. However, it is difficult if not impossible to improve without measuring and detecting where the key problems are. The PES helps pinpoint the key sources of coverage problems and can provide information regarding those aspects of the data collection that need to

be improved in future censuses, as well as estimates of undercoverage that may be used to compensate for the lost coverage. Hence, we strongly recommend that during census budgeting and planning, the required resources be set aside for a reasonable-sized PES to be carried out just after the census data-collection operation. Demographic analysis assessment of coverage is generally cheaper than a PES but it requires both access to external data sources and knowledge of demographic methods. Still, where possible, there should be budgeting for the conduct of this kind of analysis and time set aside for it as part of the main census evaluation operation.

21. In most countries, developed or not, census figures are not adjusted for undercoverage. The reason for this may be that there is no widely accepted theory or method to correct for the coverage errors, or that the reliability of undercoverage estimates from PES is not sufficient, or that political factors prevent changing of the census estimates, or the cause may be a combination of these and other factors. Hence, population estimates published from population census data remain largely without compensation for undercoverage. In some cases, information about census undercoverage, if available, may be treated as "classified" and may not be available for general user access, owing to a perception that this type of information may damage credibility of census results if inadequately interpreted. We recommend that this practice should not be adopted, but rather that results of the PES should be published or made available to relevant census user communities.

22. The above discussion relates to broad coverage of survey populations. The problem of adequate coverage evaluation is even more serious for subpopulations of special interest, such as ethnic or other minorities, because the sample size needed in a PES is generally beyond the budgetary resources available. Very little is known about how well such subpopulations are covered in censuses and other household surveys in developing countries. In Brazil, every census post-enumeration survey carried out since the 1970 census failed to provide estimates for ethnic groups or other relevant subpopulations that might be of interest. Their estimates have been limited to overall undercount for households and persons, broken down by large geographical areas (States). Results of the undercoverage estimates for the 2000 population census have recently appeared (Oliveira and others, 2003). Here we present only the results at the country level, including estimates for omission rates for households and persons for the 1991 and 2000 censuses. Undercoverage rates were similar in 1991 and 2000, with slightly smaller overall rates for 2000. One recommendation for improvement of the PES taken within Brazilian population censuses has been to expand undercoverage estimation to include relevant subpopulations, such as those defined by ethnical or age groups.

Table XI.2. Estimates of omission rates for population censuses in Brazil obtained from the 1991 and 2000 post-enumeration surveys
(Percentage)

Coverage category	1991 census	2000 census
Private occupied households	4.5	4.4
Persons living in private occupied non-missed households	4.0	2.6
Persons missed overall from private occupied households	8.3	7.9

Source: Oliveira and others (2003).

23. The figures in table XI.2 are higher than those reported for similar censuses in some developed countries. The omission rates reveal an amount of undercoverage that is non-negligible. To date, census results in Brazil are published, as is the case in the great majority of countries, without any adjustments for the estimated undercoverage. Such adjustments are made later, however, to population projections published after the census. There is a need for research to assess the potential impact of adjusting census estimates for undercoverage coupled with discussion, planning and decisions about the reliability required of PES estimates if they are to be used for this purpose.

2. Non-response

24. The term "non-response" refers to data that are missing for some survey units (unit non-response), for some survey units in one or more rounds of a panel or repeated survey (wave non-response) or even for some variables within survey units (item non-response). Non-response affects every survey, be it census or sample. It may also affect data from administrative sources that are used for statistical production. Most surveys employ some operational procedures to avoid or reduce the incidence of non-response. Non-response is more of a problem when response to the survey is not "at random" (differential non-response among important subpopulation groups) and response rates are low. If non-response is at random, its main effect is increased variance of the survey estimates due to sample size reduction. However, if survey participation (response) depends on some features and characteristics of respondents and/or interviewers, then bias is the main problem one needs to worry about, particularly for cases of larger non-response rates.

25. Särndal, Swensson and Wretman (1992, p. 575) state: "The main techniques for dealing with non-response are weighting adjustment and imputation. Weighting adjustment implies increasing the weights applied in the estimation to the y-values of the respondents to compensate for the values that are lost because of non-response ... Imputation implies the substitution of 'good' artificial values for the missing values."

26. Among the three types of non-response, unit non-response is the kind most difficult to compensate for, because there is usually very little information within survey frames and records that can be used for that purpose. The most frequent compensation method used to counter the negative effects of unit non-response is weighting adjustment, where responding units have their weights increased to account for the loss of sample units due to non-response; but even this very simple type of compensation is not always applied. Compensation for wave and item non-response is often carried out through imputation, because in such cases the non-responding units will have provided some information that may be used to guide the imputation and thus reduce bias (see Kalton, 1983; 1986).

27. Non-response has various causes. It may result from non-contact of the selected survey units, owing to such factors as the need for survey timeliness, hard-to-enumerate households and respondents' not being at home. It may also result from refusals to cooperate as well as from incapacity to respond or participate in the survey. Non-response due to refusal is often small in household surveys carried out in developing countries, mainly because, as citizen empowerment via education is less developed, potential respondents are less willing and able to refuse

cooperation with surveys; and higher illiteracy implies that most data collection is still carried out using face-to-face interviewing, as opposed to telephone interviewing or mail questionnaires. Both factors operate to reduce refusal or non-cooperation rates, and both may also lead to differential non-response within surveys, with the more educated and wealthy having a higher propensity to become survey non-respondents. At the same time, response or survey participation does not necessarily lead to greater accuracy in reporting: in many instances, higher response may actually mask deliberate misreporting of some kinds of data, particularly income- or wealth-related variables, because of distrust of government officials.

28. Population censuses in developing countries are affected by non-response. In Brazil, the population census uses two types of questionnaire: a short form, with just a few questions on demographic items (sex, age, relationship to head of household and literacy), and a larger and more detailed form, with socio-economic items (race, religion, education, labour, income, fertility, mortality, etc.), that also includes all the questions on the short form. The long form is used for households selected by a probability sample of households in every enumeration area. The sampling rate is higher (1 in 5) for small municipalities and lower (1 in 10) for the municipalities with an estimated population of 15,000 or more in the census year. Overall unit non-response in the census is very low (about 0.8 per cent in the Brazilian 2000 census). However, for the variables of the short form (those requiring response from all participating households, called the universe set), no compensation is made for non-response. There are three reasons for this: first, non-response is considered quite low; second, there is very little information about non-responding households to allow for compensation methods to be effective; third, there is no natural framework for carrying out weighting adjustment in a census context. The alternative to imputing the missing census forms by some sort of donor method is also not very popular for the first two reasons, and also because of the added prejudice against imputation when performed in cases like this. For the estimates that are obtained from the sample within the census, weighting adjustments based on calibration methods are performed that compensate partially for the unit non-response.

29. A similar approach has been adopted in some sample surveys. Two of the main household surveys in Brazil, the annual National Household Sample Survey (PNAD) and the monthly Labour Force Survey (PME), use no specific non-response compensation methods (see Bianchini and Albieri, 1998). The only adjustments to the weights of responding units are performed by calibration to the total population at the metropolitan area or State level, hence they cannot compensate for differential non-response within population groups defined by sex and age, for example. The reasons for this are mostly related to operational considerations, such as maintenance of tailor-made software used for estimation that was developed long ago and the perceived simplicity of ignoring the non-response. Both surveys record their levels of non-response, but information about this issue is not released within the publications carrying the main survey results. However, microdata files are made available from which non-response estimates can be derived, because records from non-responding units are also included in such files with appropriate codes identifying the reasons for non-response. The PME was recently redesigned (Instituto Brasileiro de Geografia e Estatística, 2002b) and started using at least a simple reweighting method to compensate for the observed unit non-response. Further developments may include the introduction of calibration estimators that will try to correct for differential non-response on age and sex. However, the relevant studies, which were motivated

by the observation that non-response is one of the probable causes of rotation group bias (Pfeffermann, Silva and Freitas, 2000) in the monthly estimates of the unemployment rate, are at an early stage.

30. A Brazilian survey that uses more advanced methods of adjustment for non-response is the Household Expenditure Survey (POF) (last round in 1995-1996, with the 2002-2003 round currently in the field). This survey uses a combination of reweighting and imputation methods to compensate for non-response (Bianchini and Albieri, 1998). Weight adjustments are carried out to compensate for unit non-response, whereas donor imputation methods are used to fill in the variables or blocks of variables for which answers are missing after data collection and edit processing. The greater attention to the treatment of non-response has been motivated by the larger non-response rates observed in this survey, when compared with the general household surveys. Larger non-response is expected given the much larger response burden imposed by the type of survey (households are visited at least twice, and are asked to keep detailed records of expenses during a two-week period). Survey methodology reports have included an analysis of non-response, but the publications presenting the main results have not.

31. Yet another survey carried out in Brazil, the Living Standards Measurement Survey (PPV), which was part of the Living Standards Measurement Study survey programme of the World Bank, used substitution of households to compensate for unit non-response. In Brazil, this practice is seldom used, and there are no other major household surveys that have adopted it.

32. After examining these various surveys carried out within the same country, a pattern emerges to the effect that there is no standard approach to compensating for, and reporting about, unit non-response. Methods and treatment for non-response vary between surveys, as a function of the non-response levels experienced, of the survey's adherence to international recommendations, and of the perceived need and capacity to implement compensation methods and procedures. One approach that could be used to improve this situation is the regular preparation of "quality profile" reports for household surveys. This might often be more practical and useful than attempting to include all available information about methods used and limitations of the data in the basic census or survey publications.

33. Regarding item non-response, the situation is not much different. In Brazilian population censuses, starting from 1980, imputation methods were used to fill in the blanks and also to replace inconsistent values detected by the editing rules specified by subject-matter specialists. In 1991 and 2000, a combination of donor methods and Fellegi-Holt methods, implemented in software like DIA (Deteccíon e Imputacíon Automática de datos) (Garcia Rubio and Criado, 1990) and NIM (New Imputation Methodology) (Poirier, Bankier and Lachance, 2001), were used to perform integrated editing and imputation of census short and long forms. In 2000, in addition to imputation of the categorical variables, imputation of the income variables was also performed, by means of regression tree methods used to find donor records from which observed income values were then used to fill in for missing income items within incomplete records. This was the first Brazilian population census in which all census records in microdata files at the end of processing have no missing values. The population census editing and imputation strategy is well documented, although most of the information regarding how much editing and

imputation was performed is available only in specialized reports. A recommendation for making access to these reports easier is their dissemination via the Internet.

34. The treatment of missing and suspicious data in other household surveys is not so well developed. In both the PNAD and the PME, computer programs are used for error detection, but there is still a lot of "manual editing", and little use is made of computer-assisted imputation methods to compensate for item non-response. If items are missing at the end of the editing phase, they are coded as "unknown". The progress made in recent years has focused on integrating editing steps with data entry, so as to reduce processing cost and time. The advent of cheaper and better portable computers has enabled IBGE to proceed towards even further integration. The revised PME for the 2000 decade started collection in October 2001 of a parallel sample, the same size as the one used in the regular survey, where data are obtained using computer-assisted (palmtop) face-to-face interviewing. There are no final reports on the performance of the palmtop computers yet, but after the first few months, the data collection was reported as running smoothly. This technology has enabled survey managers to focus on quality improvement in the source, by embedding all jump instructions and validity checks within the data-collection instrument, thus avoiding keying and other errors in the source. Non-response for income will be compensated using regression tree methods to find donors, as in the population census. However, the results of this new survey only recently became available and data collection ran in parallel with the old series for a whole year before they were released and the new series replaced the old one. A broader and more detailed assessment of the results of this new approach for data collection and processing is still under way.

35. In the PME, each household is kept in the sample for two periods of four months each, separated by eight months. Hence, in principle, data from previous complete interviews could be used to compensate for wave non-response whenever a household or household member was missed in any survey round after the first. This use of data does not occur in the old series nor is it planned for the new series, although it represents an improvement that might be considered by survey managers.

36. The pattern emerging from a cross-survey analysis of editing and imputation practices for item non-response and inconsistent or suspicious data is one of no standardization, with different surveys following different methodological paths. Censuses have clearly been the occasion for large-scale applications of automatic editing and imputation methods, with the smaller surveys not so often adopting similar methods. Perhaps there is a survey scale effect, in the sense that the investment in developing and applying acceptable methods and procedures for automatic imputation is justifiable for the censuses, but not for smaller surveys, which also have a shorter time to deliver their results. For a repeated survey like the Brazilian PME, although the time in which to deliver results is short, there would probably be a benefit to be derived from larger investment in methods for data editing and imputation because of the potential to exploit this investment over many successive survey rounds.

3. Measurement and processing errors

37. Measurement and processing errors entail observed values for survey questions and variables after data collection and processing that differ from the corresponding true values that would be obtained if ideal or gold standard measurement and processing methods were used.

38. This topic is probably the one that receives the least attention in terms of its measurement, compensation and reporting in household surveys carried out in developing and transition countries. Several modern developments can be seen as leading towards improved survey practice towards reducing measurement error. First, the use of computer-assisted methods of data collection has been responsible for reducing transcription error, in the sense that the respondent's answers are directly fed into the computer and are immediately available for editing and analysis. Also, the flow of questions is controlled by the computer and can be made to be dependent upon the answers, preventing mistakes introduced by the interviewer. The answers can be checked against expected ranges and even against previous responses from the same respondent. Suspicious or surprising data can be flagged and the interviewer asked to probe the respondent about them. Hence, in principle, data that are of better quality and less subject to measurement error may be obtained. However, there is little evidence of any quality advantages for computer-assisted interviewing over paper-and-pencil interviewing other than that of reducing the item missing-value rates and values-out-of-range rates.

39. Another line of progress has involved the development and application of generalized software for data editing and imputation (Criado and Cabria, 1990). As already mentioned in section B, population censuses have adopted automated editing and imputation software to detect and compensate for measurement error and some types of processing errors (for example, coding and keying errors), and, at the same time, item non-response. This has also occurred in some sample surveys. However, the type of compensation that is applied within this approach is capable of tackling only the so-called random errors. Systematic errors are seldom detected or compensated for using standard editing software.

40. Yet another type of development that may lead to reduction of processing errors in surveys has been the development of computer-assisted coding software, as well as data capture equipment and software.

41. Although prevention of measurement and processing errors may have experienced some progress, the same is not true of the application of methods for measuring, eventually compensating for, and reporting about measurement errors. Practice regarding measurement errors is mostly focused on prevention, and after doing what is considered important in this respect, it does not give much attention to assessment of how successful the survey planning and execution were. The lack of a standard guiding theory of measurement makes the task of setting quality goals and assessing the attainment of such goals a hard one. For example, although we do see survey sampling plans where sample size was defined with the goal of having coefficients of variation (relative standard errors) of certain key estimates below a specified value set forth in advance, we rarely see survey collection and processing plans that aim to keep item imputation levels below a specified level, or that aim at having observed measures within a specified tolerance (that is to say, maximum deviation) from corresponding "true values" with high

probability. It may be impractical to expect that realistic quantitative goals for all types of non-sampling error could be set in advance; however, we advocate that survey organizations should at least make an effort to measure non-sampling errors and use such measures to set targets for future improvement and to monitor the achievement of those targets.

C. Challenges and perspectives

42. After over 50 years of widespread dissemination of (sample) surveys as a key observation instrument in social science, the concept of sampling errors and their control, measurement and interpretation have reached a certain level of maturity despite the fact that, as we have noted, the results of many surveys around the world are published without inclusion of any sampling error estimates. Much less progress has been made regarding non-sampling errors, at least for surveys carried out in developing countries. This has not been the case by chance. The problem of non-sampling errors in surveys is a difficult one. For one thing, they come from many sources in a survey. Efforts to counter one type of error often result in increased errors of another kind. Prevention methods depend not only on technology, but also on culture and environment, making it very hard to generalize and propagate successful experiences. Compensation methods are usually complex and expensive to implement properly. Measurement and assessment are hard to perform in a context of surveys carried out under very limited budgets, with publication deadlines that are becoming tighter and tighter to satisfy the increasing demands of our information-hungry societies. In a context like this, it is correct for priority to always be given to prevention rather than measurement and compensation, but this leaves little room for assessing how successful prevention efforts were, and thereby reduces the prospects for future improvement.

43. Some users who may have poor knowledge of statistical matters may misinterpret reports about non-sampling errors in surveys. Hence, publication of reports of this kind is sometimes seen as undesirable in some survey settings mainly because of the lack of well-developed statistical literacy and culture, whose development may be particularly challenging among populations that lack broader literacy and numeracy, as is the case in many developing countries. It is also often true that statistical expertise is lacking within the producing agencies as well, leading to difficulties in recognizing the problems and taking affirmative actions to counter them, as well as in measuring how successful such actions were. In any case, we encourage the preparation and publication of such reports, with the statistical agencies striving to make them as clear as possible and accessible to literate adults.

44. Even if the scenario is not a good one, some new developments are encouraging. The recent attention given to the subject of data quality by several leading statistical agencies, statistical and survey academic associations, and even multilateral government organizations, is a welcome development. The main initiatives that we shall refer to here are the General Data Dissemination System (GDDS) and the Special Data Dissemination Standard (SDDS) of the International Monetary Fund (IMF), which are trying to promote standardization of reporting about the quality of statistical data by means of voluntary adherence of countries to either of these two initiatives. According to IMF (2001): “The GDDS is a structured process through which Fund member countries commit voluntarily to improving the quality of the data produced

and disseminated by their statistical systems over the long run to meet the needs of macroeconomic analysis." Also according to IMF: "The GDDS fosters sound statistical practices with respect to both the compilation and the dissemination of economic, financial and socio-demographic statistics. It identifies data sets that are of particular relevance for economic analysis and monitoring of social and demographic developments, and sets out objectives and recommendations relating to their development, production and dissemination. Particular attention is paid to the needs of users, which are addressed through guidelines relating to the quality and integrity of the data, and access by the public to the data." (ibid.).

45. The main contribution of these initiatives is to provide countries with: (a) a framework for data quality (see http://dsbb.imf.org/dqrsindex.htm) that helps to identify key problem areas and targets for data quality improvement; (b) the economic incentive to consider data quality improvement within a wide range of surveys and statistical output (in the form of renewing or gaining access to international capital markets); (c) a community sharing a common motivation through which they can advance the data quality discussion free from the fear of misinterpretation; and (d) technical support for evaluation and improvement programmes, when needed. This is not a universal initiative, since not every country is a member of IMF. However, 131 countries were contacted about it, and as at the present date, 46 countries have decided to adhere to the GDDS and 50 other countries have achieved the higher status of subscribers to the SDDS, having satisfied a set of tighter controls and criteria for the assessment of the quality of their statistical output.

46. A detailed discussion of the data quality standards promoted by IMF or other organizations is beyond the scope of this chapter, but readers are encouraged to pursue the matter with the references indicated here. Developing countries should join the discussion of the standards currently in place, decide whether or not to try to adhere to either of the above initiatives and, if relevant, contribute to the definition and revision of the standards. Most important, statistical agencies in developing countries can use these standards as starting points (if nothing similar is available locally) to promote greater quality awareness both among their members and staff, and within their user communities.

47. The other initiative that we shall mention here, particularly because it affects Brazil and other Latin American countries, is the Project of Statistical Cooperation of the European Union (EU) and the Southern Common Market (MERCOSUR).[25] According to the goal of the project: "The European Union and the MERCOSUR countries have signed an agreement on 'Statistical Cooperation with the MERCOSUR Countries', the main purpose of which is a rapprochement[26] in statistical methods in order to make it possible to use the various statistical data based on mutually accepted terms, in particular those referring to traded goods and services, and, generally, to any area subject to statistical measurement." The Project "is expected to achieve at the same time the standardization of statistical methods within the MERCOSUR countries as well as between them and the European Union." (For more details, visit the website: http://www.ibge.gov.br/mercosur/english/index.html). This project has already promoted a

[25] MERCOSUR is the common market of the South, a group of countries sharing a free trade agreement that includes Brazil, Argentina, Paraguay and Uruguay.

[26] The term is used here in the sense of harmonization.

number of courses and training seminars and, in doing so, is contributing towards improved survey practice and greater awareness of survey errors and their measurement.

48. Initiatives like these are essential in respect of supporting statistical agencies in developing countries to improve their position: their statistics may be of good quality, but they often do not know how good they are. International cooperation from developed towards developing countries and also between the latter is essential for progress towards better measurement and reporting about non-sampling survey errors and other aspects of survey data quality.

D. Recommendations for further reading

49. Meetings recommended as subjects for further reading include:

- International Conference on Measurement Errors in Surveys, held in Tucson, Arizona in 1990 (see Biemer and others, 1991).

- International Conference on Survey Measurement and Process Quality, held in Bristol, United Kingdom in 1995 (see Lyberg and others, 1997).

- International Conference on Survey Non-response, held in Portland, Oregon in 1999 (see Groves and others, 2001).

- International Conference on Quality in Official Statistics, held in Stockholm, Sweden in 2001 (visit http://www.q2001.scb.se/).

- Statistics Canada Symposium 2001, held in Ottawa, Canada, which focused on achieving data quality in a statistical agency from a methodological perspective (visit http://www.statcan.ca/english/conferences/symposium2001/session21/s21c.pdf).

- Fifty-third session of the International Statistical Institute (ISI), held in Seoul, Republic of Korea in 2001, where there was an invited paper meeting on “Quality programs in statistical agencies”, dealing with approaches to data quality by national and international statistical offices (visit http://www.nso.go.kr/isi2001).

- Statistical Quality Seminar 2000, sponsored by IMF, held in Jeju Island, Republic of Korea in 2000 (visit http://www.nso.go.kr/sqs2000/sqs12.htm).

- International Conference on Improving Surveys, held in Copenhagen, Denmark in 2002 (visit http://www.icis.dk/).

References

Bianchini, Z.M., and S. Albieri (1998). A review of major household sample survey designs used in Brazil. In *Proceedings of the International Conference on Statistics for Economic and Social Development*. Aguascalientes, Mexico, 1998: Instituto Nacional de Estadística, Geografía e Informática (INEGI).

Biemer, P.P., and R.S. Fecso (1995). Evaluating and controlling measurement error in business surveys, Cox and others, eds. In *Business Survey Methods*, New York: John Wiley and Sons.

Biemer, P.P., and others (1991). *Measurement Errors in Surveys*. New York: John Wiley and Sons.

Correa, S.T., P.L. do Nascimento Silva and M.P.S. Freitas (2002). Estimação de variância para o estimador da diferença entre duas taxas na pesquisa mensal de emprego. In *15º Simpósio Nacional de Probabilidade e Estatística.* Aguas de Lindóia, Brazil, São Paulo, Brazil: Associação Brasileira de Estatística.

Criado, I.V., and M.S.B. Cabria (1990). *Procedimiento de depuración de datos estadísticos,* cuaderno 20. Vitoria-Gasteiz, Spain: EUSTAT Instituto Vasco de Estadística.

Garcia Rubio, E., and I.V. Criado (1990). DIA System: software for the automatic imputation of qualitative data. In *Proceedings of the United States Census Bureau Sixth Annual Research Conference* (Arlington, Virginia). Washington, D.C.: United States Bureau of the Census.

Groves, R.M., and others (2001). *Survey Non-response*. New York: John Wiley and Sons.

Instituto Brasileiro de Geografía e Estatística (2002a). http://www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad99/metodologia99.shtm.

__________ (2002b). http://www.ibge.net/home/estatistica/indicadores/trabalhoerendimento/pme/default.shtm.

International Monetary Fund (2001). Guide to the General Data Dissemination System (GDDS). Washington, D.C.: IMF Statistics Department. Available from http://dsbb/imf/org/applications/web/gdds/gddsguidelangs).

Kalton, G. (1983). *Compensating for Missing Survey Data*. Research Report Series. Ann Arbor, Michigan: Institute for Social Research, University of Michigan.

__________ (1986). Handling wave non-response in panel surveys. *Journal of Official Statistics*, vol. 2, No. 3, pp. 303-314.

__________ (2002). Models in the practice of survey sampling (revisited). *Journal of Official Statistics*, vol.18, No. 2, pp. 129-154.

Kordos, J. (2002). Personal communications.

Lyberg, L., and others, eds. (1997). *Survey Measurement and Process Quality*. New York: John Wiley and Sons.

Marks, E.S. (1973). The role of dual system estimation in census evaluation. Internal report. Washington, D.C.: United States Bureau of the Census.

Oliveira, L.C., and others (2003). *Censo Demográfico 2000: Resultados da Pesquisa de Avaliação da Cobertura da Coleta*. Textos para Discussão, No. 9. Rio de Janeiro: IBGE, Directoria de Pesquisas.

Pfeffermann, D., P.L. Nascimento de Silva and M.P.S. Freitas (2000). Implications of the Brazilian Labour Force rotation scheme on the quality of published estimates. Internal report. Rio de Janeiro: IBGE, Departamento de Metodologia.

Platek, R., and C.E. Särndal (2001). Can a statistician deliver? *Journal of Official Statistics*, vol. 17, No. 1, pp. 1-20.

Poirier, P., M. Bankier and M. Lachance (2001). Efficient methodology within the Canadian Census Edit and Imputation System (CANCEIS). Paper presented at the Joint Statistical Meetings, American Statistical Association.

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

United Nations (1982). *National Household Survey Capability Programme: Non-sampling errors in household surveys: sources, assessment and control*: *Preliminary Version*. DP/UN/INT-81-041/2. New York: Department of Technical Cooperation for Development and Statistical Office.