Proceedings of Statistics Canada Symposium 2001
Achieving Data Quality in a Statistical Agency: A Methodological Perspective

HOW IMPORTANT IS ACCURACY?

Graham Kalton[1]

ABSTRACT

Accuracy is only one of the dimensions in the broader context of quality that has been articulated in recent years. Yet it remains of critical importance. Inaccuracy in survey estimates can arise from sampling error, nonresponse, noncoverage, measurement error, and processing error. It is argued that research is needed on all these components of accuracy and that the results should be communicated to users. Suppression of seriously inaccurate estimates should be undertaken with care. Users should have a say in determining the fitness for use of an estimate. Users should be trained to be able to assess information on accuracy in an informed way.

KEY WORDS: Quality dimensions; Sources of inaccuracy; Fitness for use.

1. DIMENSIONS OF QUALITY

The longstanding tradition in the survey literature has been to focus on the accuracy of the survey results, where accuracy is generally expressed in terms of low levels of sampling and nonsampling errors. The concept of total survey error has been advanced, with the idea that surveys should be designed to minimize the total error in the survey estimates by allocating resources to achieve the optimum balance of various error sources. Yet those designing statistical data collections have always recognized that aspects other than accuracy need to be considered. For example, a choice must sometimes be made between measuring a desired concept directly but inaccurately and measuring a related concept accurately. This is the issue of relevance. As another example, estimates based on somewhat imperfect provisional data can sometimes be produced appreciably sooner than estimates from final data. This leads to the common practice of revisions in economic data. This is the issue of timeliness.

Despite the widespread recognition that quality of statistical data involves more than just accuracy, the acknowledgement of this fact has entered the literature only in the past few years. Thus, for example, Brackstone (1999) identifies the quality dimensions of relevance, accuracy, timeliness, accessibility, interpretability, and coherence. Arondel and Depoutot (1998) and Grünwald and Linden at this symposium provide a similar list of dimensions. The articulation of these dimensions is an important contribution to the field since it provides a framework for the overall assessment of quality. The explicit listing of the other dimensions in addition to accuracy undoubtedly makes them more salient, and thus may draw greater attention to them. Yet accuracy remains of critical importance.

The dimensions of quality given above are those that should be taken into account in developing statistical programs in national statistical offices to best serve user needs. In a discussion of a paper by Platek and Särndal (2001), Fellegi (2001) identifies two issues that he terms “survival issues beyond quality”. One is respect for respondents, in which he includes privacy, confidentiality (including security), and management of respondent burden. The other is “credibility of information”, which includes accuracy, transparency, nonpolitical objectivity, and relevance. Note that accuracy occurs here as a survival issue, a point to which I will return later.

The quality dimensions are not all-or-none properties, but matters of degree. There are often trade-offs to be made between them (Holt and Jones, 1998). For instance, an improvement of a certain amount in

[1] Graham Kalton, Westat, 1650 Research Blvd, Rockville, Maryland, U.S.A. 20850.

timeliness may be bought at the price of lower accuracy. The optimum choice of design and procedures for a statistical data collection system thus involves an assessment of the properties of the alternatives along all dimensions, and an effective balancing of them. At least a minimum degree of quality needs to be achieved on each of the quality dimensions before a study goes ahead.

2. DIMENSIONS OF ACCURACY

In the same way that quality is multidimensional, with accuracy as one of its dimensions, so accuracy is also multidimensional. The main dimensions of accuracy, or rather inaccuracy, are sampling error, nonresponse, noncoverage, measurement error, processing error, and computational error. The attention paid to these various sources of error in designing surveys and evaluating their estimates is generally geared more to what can be easily measured and reported than what is most important. Thus, sampling errors and nonresponse receive a good deal of attention, whereas the other sources of error are often largely ignored. In their paper at this symposium, Grünwald and Linden describe the difficulty that Eurostat has in obtaining information on the other sources of error for surveys in the European Union.

Nonresponse rates are usually discussed mainly in terms of total nonresponse. Much less attention is given to item nonresponse, perhaps because of the complexity of assessing the multitude of rates for all the survey items. Yet, in some surveys and for some items, item nonresponse can be a much more serious problem than total nonresponse. The primary concern with missing data from either total or item nonresponse is the risk that they may cause bias in the survey estimates. However, information on nonresponse bias is rarely available, and this is particularly the case after weighting adjustments and imputation have been used to compensate for the missing data. Nonresponse rates are easy to produce, but they serve only as indicators of the potential for bias and of the quality of the data collection procedures. More attention could usefully be given to nonresponse bias.

Another source of missing data – noncoverage – generally receives little attention, probably because it is less easy to measure. Yet noncoverage can be a greater concern than nonresponse, particularly for some population subgroups. The 1996 U.S. Current Population Survey, for instance, had a response rate of 93.4 percent, an overall coverage ratio of 92.5 percent, but a coverage ratio of only 66.2 percent for black males aged 20-29 years old (U.S. Census Bureau and U.S. Bureau of Labor Statistics, 2000). The paper by Post and Walker at this symposium, discussing the coverage errors in the Canadian Labour Force Survey, is a welcome contribution on this neglected error source.

Measurement errors are often a major cause of inaccuracy in survey estimates, but they are hard to measure and hence are rarely quantified. Special studies are needed to investigate the validity or reliability of survey responses. Since such studies generally require the investment of substantial resources, they are more likely to be carried out in connection with repeated surveys, when the study results may lead to improvements in future rounds of the survey, rather than in connection with ad hoc surveys.

Processing errors can enter at various stages of processing the data for analysis. They can arise during data entry, editing and coding. Although errors in coding complex phenomena such as industry and occupation can be substantial and can be readily investigated with reliability studies, such studies are seldom conducted. More attention could be given to this error source.

A full evaluation of the accuracy of a survey estimate requires an assessment of the magnitude of all the sources of error listed above. It is rare that all the information is available to make such an evaluation, but there does appear to be a growing interest in assessing the full range of error sources. It is important to measure all the error sources in order to target improvements needed in future surveys and also to fully inform users about the accuracy of the survey estimates. For the latter purpose, effective means of communicating findings on accuracy to users are needed. The report of the U.S. Federal Committee on Statistical Methodology's Subcommittee on Measuring and Reporting the Quality of Survey Data (2001) provides useful guidance on these issues.

3. FITNESS FOR USE

While accuracy is an important property of survey estimates, it cannot be treated in the abstract. It is closely interrelated with other factors, such as the available budget, the need for timely results, the need for relevant estimates, and the need to control response burden. In evaluating the fitness for use of a survey estimate, a user needs to assess all these elements to determine whether the estimate is acceptable for the purpose in hand. An important consideration is that the user should have ready access to the information needed to be able to make at least a general assessment of the level of accuracy. The user also needs to have the skills to make that assessment.

An issue of debate is whether a statistical office should make its own judgement of a minimum acceptable level of accuracy and suppress estimates that fail to reach that level. Many statistical offices, for instance, suppress estimates with coefficients of variation (CVs) of greater than, say, 30 percent. Is this appropriate? Similarly, should model-dependent estimates for small domains be suppressed if they fail to meet a minimum level of accuracy? Also, should the data from a survey that achieved a response rate of only, say, 40 percent be suppressed?

From one perspective, it can be argued that estimates of low accuracy should not be suppressed since they add something to current knowledge. Users may be better served by having estimates that are potentially very inaccurate rather than having nothing at all. If estimates of very low accuracy are provided, they must, of course, be accompanied by firm warnings about their limitations.

As a specific example, consider the model-dependent estimates of poor school age children in school districts that the U.S. Census Bureau produces for the U.S. Department of Education for use in allocating funds to school districts to support programs for educationally disadvantaged children. It could well be argued that these estimates should be withheld because of their high level of inaccuracy. However, a panel of the Committee on National Statistics of the National Research Council recommended the use of these estimates for the fund allocation on the grounds that they were at least as good as the alternatives available (Citro and Kalton, 2000). This example brings out the point that a user's judgement of fitness for use need not be based on a set standard but may be dependent on the alternatives available.

One argument for suppression is that many users lack the training needed to properly assess warnings of low accuracy, or they are too cavalier in heeding these warnings. They should therefore be protected by not being given estimates of low quality. Another argument is that a statistical office should not produce low quality estimates because they will tarnish the reputation of all the estimates that the office produces. This point relates to Felligi's survival issue of credibility of information mentioned earlier.

There are thus good arguments for and against suppression. I favor providing the estimates in most cases, with suitable warnings, and with extensive ongoing efforts to educate users to appreciate the warnings. A distinction can be made here between estimates whose high level of inaccuracy can be measured (e.g., by a CV) and estimates whose potential for inaccuracy can only be inferred (e.g., from a low response rate). I am more comfortable with publishing the estimates in the former case than in the latter. However, suppose a survey achieves a very low response rate on a topic for which there are no alternative, reasonably current, data. In such a case, I believe that strenuous efforts should be made to attempt to counteract potential biases (e.g., by multivariate weighting adjustments), and that the survey should then be analyzed and the results reported with due warnings.

4. CONCLUDING REMARKS

In summary, I believe that accuracy remains a critically important component of overall quality. Increased research efforts are needed to investigate the various sources of inaccuracy, both to help guide improvements in survey methods and to fully inform users about the overall accuracy of the estimates they are using. User needs should affect decisions about fitness for use. Producers should make greater efforts to communicate data on accuracy to users and to encourage them to use those data. Training programs for users could be a valuable part of these efforts.

REFERENCES

Arondel, P. and Depoutot, R. (1998), “Overview of Quality Issues when Dealing with Socio-Economic Products in the International Environment”, paper presented at the XXXth ASU meeting.

Brackstone, G. (1999), “Managing Data Quality in a Statistical Agency,” *Survey Methodology*, 25, pp. 139-149.

Citro, C.F. and Kalton, G. (eds.) (2000), *Small-Area Estimates of School-Age Children in Poverty. Evaluation of Current Methodology*, Panel on Estimates of Poverty for Small Geographic Areas, Committee on National Statistics, National Research Council, Washington, D.C.: National Academy Press.

Fellegi, I.P. (2001), “Comment”, *Journal of Official Statistics*, 17, pp. 43-50.

Holt, D. and Jones, T. (1998), “Quality Work and Conflicting Quality Objectives”, paper presented at the 84th DGINS conference, Stockholm, Sweden. http://www.statistics.gov.uk/ methods_quality/.

Platek, R. and Särndal, C.-E. (2001), “Can a Statistician Deliver?”, *Journal of Official Statistics*, 17, pp.1-21.

U.S. Census Bureau and U.S. Bureau of Labor Statistics (2000), *Current Population Survey. Design and Methodology*, Technical Paper 63, Washington D.C.: U.S. Census Bureau and U.S. Bureau of Labor Statistics.

U.S. Federal Committee on Statistical Methodology, Subcommittee on Measuring and Reporting the Quality of Survey Data (2001), *Measuring and Reporting Sources of Error in Surveys*, Statistical Policy Working Paper 31, Washington, D.C.: Statistical Policy Office, Office of Management and Budget.