The Use of Paradata in Survey Research

Mick Couper
Institute for Social Research
University of Michigan
Ann Arbor, MI 48106
Mcouper@umich.edu

Lars Lyberg
Statistics Sweden
Box 24300
10451 Stockholm
Sweden
Lars.Lyberg@scb.se

1. Introduction

In recent years a literature on process quality has emerged (Lyberg et al. 1997, Couper 1998, Scheuren 2001a, Morganstein and Marker 1997, Heerwegh 2002, Groves and Heeringa, 2004, and Biemer and Caspar 1994). This literature has adopted the following view: Quality can be viewed as a three-level concept. The first level, product quality, is the set of product characteristics ideally established with the main users. The second level, process quality, decides the product quality. If the process is error-free, so is the product that is the result of the process. The third level, organizational quality, decides the process quality. If organizational characteristics such as leadership and competence are not up to par, then the organization is incapable of developing dependable processes. The three levels are intertwined but the process quality is at the core. Typically process quality is evaluated and controlled by observing key process variables. In Couper's (1998) terminology these observations are called paradata (data about the process) in contrast to related concepts such as data (about the phenomena under study), and metadata (data about the data).

Process quality and paradata are of course not new. What is new is the structured approach in choosing, measuring and analyzing key process variables and that quality can be viewed on different levels. In this paper we discuss the importance of paradata in survey research. We discuss some terminology and the fact that users have very limited influence on process management and collection of paradata. We provide a few examples of process management and identify some potential lines of development.

2. Some Terminology

Paradata is part of the large family of survey quality concepts. Scheuren (2001b) distinguishes between macro and micro paradata. Typical macro paradata include global process measures such as response rates and coverage rates. As the name suggests micro paradata are process data that concern individual records. One example is the practice of flagging imputed records. Micro paradata can also be at a lower level, such as item-level time-stamps or even keystroke data (e.g., Couper, Hansen, and Sadosky, 1997). An overwhelming portion of paradata are in macro form and often part of or a byproduct from quality control operations, such as process control, where paradata might be plotted on control charts to check if variation tends to be natural or has its origin in special causes. With the widespread adoption of computer assistance in survey data collection (CATI, CAPI, Web, etc.), paradata are increasingly produced at lower levels of detail as an automatic byproduct of the process itself. Paradata might also come from quality control operations using acceptance sampling where the paradata might be the number of defects in batches of units, where the purpose is to accept or reject batches of work based on number of defects found. There are theories for these methods called statistical quality control theories (Montgomery 2005). Statistical quality control theory has found applications in survey operations that resemble those in a factory assembly line, such as keying, coding and printing (for reviews see Biemer & Lyberg 2003; Mudryk et al. 2001). For other survey processes the notion of paradata is much looser. During the last 20 years, however, with the advent of TQM and other quality management systems the control chart has seen a revival in survey work. Morganstein and Marker (1997) and Deming (1986) advocate the use of the control chart as a tool for distinguishing between different kinds of variation and for developing stable and predictable processes. Their line of thought is not confined to just survey processes, but rather to all processes of importance to the statistical organization. Examples of such processes are budgeting, recruiting, and training. So the paradata concept is important for achieving organizational quality as well.

3. The Importance of Paradata

Paradata are thus collected to shed some light on the survey processes that lead to the survey products or services. Paradata can have many specific uses, though. First, they can be used as a basis for continuous updates of progress. Managers get materials such as control charts or indicator reports containing information such as nonresponse rates, productivity, edit failure rates by edit check, interviewer monitoring outputs, number of coding cases that go to adjudication, etc. depending on what processes are studied. Based on these reports managers can act in various ways or choose not to act. The continuous element could mean anything from daily checks (once or several times a day) initiated by the system to checks initiated by the managers themselves. Basically managers use these reports to make sure that processes are running the way they are supposed to. It is of course important that managers have been trained in analyzing the paradata outcomes. One obvious case is when paradata shows that a process adjustment is called for. Then paradata function as an early warning system.

Second, paradata can be used as an input to long-run process improvement. Control charts may not reveal any variation considered unnatural but the variation is still not acceptable from a management point of view. Thus a process improvement project might be triggered by the paradata.

Third, paradata may suggest methodological changes. For instance, one type of paradata are keystroke files secured from computer-assisted interviewing. When analyzing these files it turns out that for some specific questions interviewers have to go back frequently and change the responses. In editing paradata might show that edit changes are concentrated to a few questions. These are examples of situations where root causes of problems can be eliminated (with or without experimentation). Another example is that paradata generate rich timing and item missing data information on an item-level that can be used to make changes in the instrument or process as the survey progresses.

Fourth, Groves and Heeringa (2004) have discussed the use of paradata as part of what they calls “responsive designs.” Here paradata are used to inform cost and quality tradeoff decisions in real time. The ability to monitor both paradata and regular survey data creates an opportunity to alter the design during the survey data collection in order to improve survey cost efficiency and achieve more accurate estimates.

Fifth, obviously some paradata analyses might very well be used to upgrade the organizational quality. For instance, paradata on editing might suggest that editing be performed in a centralized fashion rather than survey by survey.

4. Producers and Users

As pointed out by Scheuren (2001b), there is a tendency for paradata to be producer-oriented. The global measures mentioned, such as nonresponse rates or coding error rates, are clearly insufficient from a user perspective. They can successfully be used by the producer to develop better processes but they tell the user almost nothing. The survey world is full of these global indicators that tell very little about the real survey quality.

There are two things the producers can do. First, they can use paradata and development work to reduce errors in the survey processes. Second, as Scheuren points out, the statistical producers must upgrade their user communication capabilities. Very few users have knowledge of or ideas on how to handle information on, say, nonsampling errors. Communication with users is basically one-way (U.S. Federal Committee on Statistical Methodology 2001) as documented in various report studies in the U.S. The situation is hardly any better in other countries. Typically these quality reports treat the impact of errors more or less verbally and quality reports are "one size fits all." Thus there is need for new thinking when it comes to quality assessments that could be used by customers.

5. Some Examples of Paradata

Morganstein and Marker (1997) have developed a structured approach toward the selection, measurement and analysis of what the call key process variables. In their approach they define a set of product characteristics such as low nonresponse bias, a certain release date, and maximum cost. Associated with these characteristics are a number of underlying processes. These processes are checked by means of key process variables. For a process variable to be key it has to have a distinct influence on the product characteristic. For instance, to achieve a low nonresponse bias it is probably more important to get good cooperation from specific groups than getting a good overall response rate. The choice of key process variables is far from trivial. We might have to go through several steps before we have a set of variables that can serve as variables that have a distinct effect on the desired process outcome. The choice can usually be guided by using cause-and-effect diagrams where likely process variables are listed and evaluated.

The Leadership Group on Quality (Lyberg et al. 2001) emphasized the need for paradata in the European Statistical System. Recently, Aitken et al. (2004) have developed a handbook on improving quality by analysis of process variables. The handbook contains numerous examples of the use of paradata for various processes. Other examples include control charts for interviewers (Mudryk et al, 1996; Statistics Sweden 2004), and control charts in transportation (Pierchala and Surti 1999). It should be pointed out that paradata are not only the outcome of formal control charts. Couper (1998) lists a large set of paradata that can be used in a CASIC environment including behavior coding, usability testing, trace file analysis, monitoring, reinterviews, production reports, time measures, tracking data, problem reports, and interviewer notes. Obviously some of these measures are not key process variables in the sense that they are measured continuously. They may be collected once per data collection effort but they are still paradata in the sense that they help shed light on the process.

6. Some Thoughts for Development

The ideal situation is that the process variables that we choose are really key so that by measuring and analyzing them we get a chance to check or adjust the process so that the product is delivered according to specifications. Reality is more complicated. The key process variables are not always key. They are merely indicators with limited prediction value. One example is the total response rate that usually has a limited value as a predictor of nonresponse bias. The development of alternative indicators of nonresponse error is one important area for research.

New technology permits the generation of massive amounts of paradata. One might talk about paradata databases. Furthermore paradata are multivariate in nature. There are situations where paradata need to be combined to be relevant and any process changes have to be considered as trade-offs. Therefore we need more practical methods for process management. We need to learn how to make use of the paradata, not only to understand the process better and evaluate the quality of the process after the fact, but also to use the indicators in an efficient way to intervene in the process as needed.

7. References

Aitken et al. (2004), *Handbook on Improving Quality by Analysis of Process Variables*. Eurostat

Biemer, P.P., and Caspar, R.A. (1994), “Continuous Quality Improvement for Survey Operations: Some General Principles and Applications.” *Journal of Official Statistics*, 10: 307-326.

Biemer, P.P., and Lyberg, L.E. (2003), *Introduction to Survey Quality*. New York: Wiley.

Couper, M.P. (1998), “Measuring Survey Quality in a CASIC Environment.” Paper presented at the Joint Statistical Meetings of the American Statistical Association, Dallas, August.

Couper, M.P., Hansen, S.E., and Sadosky, S.A. (1997), “Evaluating Interviewer Performance in a CAPI Survey.” In L. Lyberg *et al.* (eds.), *Survey Measurement and Process Quality*. New York: Wiley.

Deming, W.E. (1986), *Out of the Crisis*. Cambridge: Cambridge University Press.

Groves, R.M., and Heeringa, S.G. (2004), “Responsive Design for Household Surveys: Tools for Actively Controlling Survey Nonresponse and Costs.” University of Michigan, unpublished paper.

Heerwegh, D. (2003), “Explaining Response Latencies and Changing Answers Using Client-Side Paradata from a Web Survey.” *Social Science Computer Review*, 21 (3): 360-373.

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

Lyberg, L.E. et al. (2001), ”Report from the Leadership Group on Quality.” Paper presented at the International Conference on Quality in Official Statistics, May, Stockholm, Sweden.

Montgomery, D. (2005), *Introduction to Statistical Quality Control*. New York:Wiley

Morganstein, D.R., and Marker, D.A. (1997), “Continuous Quality Improvement in Statistical Agencies.” In L. Lyberg *et al.* (eds.), *Survey Measurement and Process Quality*. New York: Wiley.

Mudryk, W., Burgess, M.J., and Xiao, P. (1996), “Quality Control of CATI Operations in Statistics Canada.” Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 150-159.

Mudryk, W., Bougie, B., Xiao, P., and Yeung, A. (2001), *Statistical Methods in Quality Control at Statistics Canada*. Course Reference Materials, Statistics Canada, Ottawa.

Pierchala, C. and Surti, J. (1999), “Control Charts as a Tool in Data Quality Improvement.” NHTSA Technical Report, U.S. Department of Transportation.

Scheuren, F. (2001a), “How Important is Accuracy?” *Proceedings of Statistics Canada Symposium*. Ottawa: Statistics Canada.

Scheuren, F. (2001b), “Macro and Micro Paradata for Survey Assessment.” Washington, D.C.: Urban Institute, unpublished paper.

Statistics Sweden (2004), “Quality Assurance of Interview Work at Statistics Sweden.” Project Report SÄK, Statistics Sweden (in Swedish).

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