

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE**

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Vienna, Austria, 21-23 April 2008)

Topic (i): Editing of data acquired through electronic data collection

TRYING TO IMPROVE EDITING TASKS THROUGH EDR METHODS

Supporting Paper

Prepared by Pedro Revilla, Ignacio Arbués, Margarita González and Isabel Yun
INE, Spain

I. INTRODUCTION

1. Data editing is a crucial part of surveys, because it directly affects data quality. Moreover, it is an expensive part of the statistical production process. Recently, more and more statistical agencies are researching new methods and strategies in order to improve the efficiency of traditional data editing. Many different methods are being used in order to meet this target. For example, several generalized methods are being proposed. Selective editing approaches try to improve the efficiency of the editing process, selecting only a subset of the questionnaires to be edited manually. Although years ago it was considered customary to edit manually every questionnaire, today it is considered inefficient because most of the editing work has no impact at the aggregate level. Different methods of selective editing are being used in many countries and extensive research is being conducted.

2. Electronic Data Reporting (EDR) methods offer new opportunities for improving editing tasks. In mail surveys data editing has typically been carried out after data collection. However, EDR methods allow data editing to be carried out interactively, during data collection. Moving editing closer to respondents can significantly contribute to improving editing effectiveness. Hence, EDR offers the opportunity for re-engineering statistical production processes, in a way that allows reporting enterprises to play a more active role in data editing and provide data of higher quality.

3. Like many other statistical agencies, INE has a significant interest in EDR methods. Concerning business surveys, Web questionnaires are offered as a voluntary option to fill in the questionnaires. Because the reduction in the enterprise burden using Web questionnaires is not always obvious, we have found that the take-up of electronic data reporting is quite low. For this reason, we try to encourage the use of Web questionnaires. A key success factor in encouraging the use of Web questionnaires is giving enterprises some incentives such as free deliveries of tailored data. When an enterprise sends a form, it immediately receives tailored data from the server. Offering these data through the Web has some advantages over sending this same data on paper by mail. Taking these advantages into account, we expect more enterprises to use the Web survey.

4. We have established an “auditing system”, carried out by the reporting enterprises (Revilla, 2005). This auditing system comes from the Total Quality Management (TQM) action of offering to the enterprises free tailored data in order to improved our relationship with them. EDR offers new opportunities on this issue.

5. The outline of this paper is as follows. In Section II we discuss the opportunities associated with EDR methods. In Section III we present an “auditing system” carried out by the reporting enterprises.

Section IV describes how we try to encourage reporting enterprises to use EDR. The paper ends with some final remarks.

II OPPORTUNITIES ASSOCIATED WITH EDR METHODS

6. EDR methods offer new opportunities for improving editing tasks and getting high quality incoming data. Whereas Computer Assisted Interviewing integrates into one stage previously distinct phases such as interviewing, data capture and editing, Computerized Self-Administered Questionnaires (CSAQs) go a step further by shifting such activities to the respondent.

7. Several advantages could be expected from using CSAQs questionnaires. Improving accuracy results from built-in edits, which allow the reporting enterprises to avoid errors as they are made. The elimination of data keying at the statistical agency directly gets rid of a common source of error. Some electronic devices (automatic data fills and calculations, automatic skipping of no applicable questions, etc.) could help the respondent to fill in the questionnaire easier and faster. The cost for statistical offices to carry out a survey using the CSAQs questionnaires could decrease. Savings could be achieved from reducing storage, packing, postal charges and eliminating data keying and keying verification. Some of the editing task could be reduced from built-in edits. Although an improvement on data quality could be expected from CSAQs questionnaires, it is very difficult to measure the real impact on data quality, given the self-selective nature of the respondents that choose the electronic option.

8. For most of the surveys, EDR cannot be at the moment the only way of data collection. Paper data collection is probably going to stay with us for some years. Hence, a mixed mode of data collection (partly paper, partly electronic) should be used. Global strategies should be designed, because data editing strategies differ whether using paper or an electronic questionnaire.

9. Many statistical offices are experimenting with the use of different EDR options in data collection (CATI, CAPI, XBRL, etc.). Web questionnaires offer some advantages over other more complex EDR methods. The Web is a mature technology for EDR because of widespread public acceptance in enterprises and institutions. The prerequisites are only a PC, access to the Internet, and a browser. There is no need, in principle, to incorporate other software on the reporting enterprises. The Web makes it simple to put electronic forms at the disposal of almost every enterprise, whatever its size.

10. Concerning the edits to be implemented, some crucial questions arise: What kind of edits should be implemented on the CSAQs questionnaires? How many? Only fatal edits or fatal edits and query edits? What kind of edits should be mandatory? On one hand, we need to include some edits. If we do not, then the information collected by CSAQs questionnaires should be treated to the editing procedures in exactly the same way as collected by paper. In that case, we would lose an essential advantage of CSAQs questionnaires: no need to editing again the information with a suitable set of edits implemented in the CSAQs application. On the other hand, we need to be extremely careful in the set of edits to be implemented, because if we implement a big set, then respondents will give up and prefer the freedom they have in paper. Too many edits could even irritate the reporting enterprises and increase the burden. In that case we will lose all the advantages of CSAQs questionnaires, as users will prefer the easy way (paper).

III. AUDITING BY REPORTING ENTERPRISES

11. In the National Statistical Institute of Spain, we have established an “auditing system”, carried out by the reporting enterprises. This auditing system comes from the TQM action of offering to the enterprises tailored data on their market shares as a payment for filling in the questionnaires (Gonzalez and Revilla, 2002).

12. When we implemented this action we were not entirely aware of its potential use as an auditing system. We only tried to improve our relation with reporting enterprises, offering tailored data on their

market share as a “payment” for filling in the questionnaires. The frequent complains of enterprises about statistical burden was the origin of our offer.

13. Hence, we offer tailored data on market share, in exchange for the questionnaires. This data are computed based on the data reported by each enterprise, comparing them with those of the rest of enterprises. We provide the reporting enterprises with answers to the following questions: What is my market share in my business activity?, How many enterprises have a larger market share than mine?, What is the overall share of those enterprises with a larger market share than mine?

14 The reaction of enterprises to this offer has been quite positive. Hence, we are trying to use a new model of relationship with enterprises, that we have named the “joint venture model”. The underlying principle of this model is that our relationship with reporting enterprises should be based on mutual use and collaboration, rather than on the legal duty of the enterprises to fill in compulsory questionnaires.

15. We introduced this action for the first time in the 1992 Annual Industrial Survey. Since then, it came to our knowledge the additional use of this action as an external evaluation method. Some enterprises began to call us, when our data on market shares did not match their expectations. Different errors could be detected and corrected from these conversations. In others cases, some questions and underlying definitions could be improved. We then entirely realised the utility of that feedback from reporting enterprises.

16. Hence, we tried to encourage that feedback as much as possible. Currently, more than 11000 enterprises have received our market studies as a payment for filling in the questionnaires. Hence, it would be possible to say that we have 11000 external auditors, who evaluate the quality of our statistical products. Nevertheless, as we have a sample of about 70000 enterprises, only one in six reporting enterprises have participated in this action. What we have noticed is that the person who fill in the questionnaire, is not usually the person inside the enterprise who would be interested in our market studies. We have to make a lot of efforts in order to inform the enterprises’ management of our data offer.

17. Web surveys offer new opportunities on this issue. Working in this direction is the last step in our TQM project of giving reporting enterprises free of charge tailored data in exchange for the questionnaires (Arbués et al., 2006).

IV. ENCOURAGING REPORTING ENTERPRISES TO USE EDR

18. INE has a significant interest in EDR methods. An example of this was the possibility offered to all citizens to fill in the Population Census 2001 using the Internet. The next Agricultural Census (2009) will use a CAPI system for first time. Concerning business surveys, Web questionnaires are offered as a voluntary option to fill in the questionnaires. A major target of this project is offering the respondents another option to fill in the questionnaires, in the hope of reducing respondent burden, or, at least, improving our relationship with them.

19. Because the reduction in the enterprise burden using Web questionnaires is not always obvious, many respondents have not changed to the Web form. We have found that the take-up of electronic data reporting for enterprise surveys is generally less than 25%, and often less than 15%. The take-up for household surveys is even lower. In the case of the Population Census 2001 it was less than 1%! Studies in other countries also find low rates of response via Internet when a mixed mode is used. Even though the usage of the electronic option by respondents has increased lately (for example, Paula Weir, 2005) it still leaves room for improvement.

20. For this reason, we try to encourage the use of Web questionnaires. Working in this direction is the last step in our TQM project of giving reporting enterprises free of charge tailored data in exchange for the questionnaires (Arbués et al., 2006). When an enterprise sends a valid form (i.e. passing the mandatory edits), it immediately receives tailored data from the server. Offering this data through the

Web has some advantages (speed, possibility to edit the file) over sending this same data on paper by mail. Taking these advantages into account, we expect more enterprises to use the Web survey.

21. We are testing this action in the Turnover and New Orders Survey. This monthly survey uses a very simple form that includes only two variables: turnover and new orders, broken down by geographic markets (14 fields are requested). The sample size is of about 13.500 establishments. The editing strategy (built-in edits, selective editing and macroediting) is described in Arbués et al. (2006).

22. The basic idea is that, after Web edits, no traditional microediting is needed. A selective editing approach is used in a way that the most influential suspicious values could be detected. Hence, all fatal errors and the most important query errors can be corrected before the index is disseminated for the first time.

23. The tailored data that we sent to the enterprise consist of graphs showing the enterprise trend and its position in relation with its sector. Some examples of these data can be seen in Figures 1 (new orders annual rate for the establishment and its sector) and 2 (turnover annual rate).

V. FINAL REMARKS

24. EDR methods offer many opportunities for improving editing tasks. But probably, the electronic questionnaires have not the same advantages for the reporting enterprises than for the statistical offices. The respondents' benefits need to be clearly explained to convince them to use the Web questionnaire. An important element to improve the acceptance among reporting enterprises is to consider Web questionnaires in a wider context of all their administrative duties and of all electronic data reporting. It is unlikely that reporting enterprises are willing to adapt their systems only for statistical purposes.

25. Hence, statistical offices should be aware of the habits of respondents and try to adapt electronic questionnaires to these trends (for example, e-commerce, e-administration, XBRL, etc.). A key success factor in encouraging the use of Web questionnaires is giving enterprises some incentives such as temporary access to information or free deliveries of tailored data.

VI REFERENCES

Gonzalez, M. and Revilla, P. (2002). "*Encouraging respondents in Spain*". The Statistic Newsletter. OECD. Issue No. 12.

Revilla, P. (2005), "*Auditing by reporting enterprises*". 55th Session of the International Statistical Institute. Sydney. April, 2005

Weir, P. (2005), "*The Movement to Electronic Reporting and the Impact on Editing*". 55th Session of the International Statistical Institute. Sydney. April, 2005

Arbués, I., Gonzalez-Villa, M and Revilla, P.(2006), "Editing Web Questionnaires: Challenges and Opportunities". UN/ECE Work Session on Statistical Data Editing. Bonn. Sep. 2006

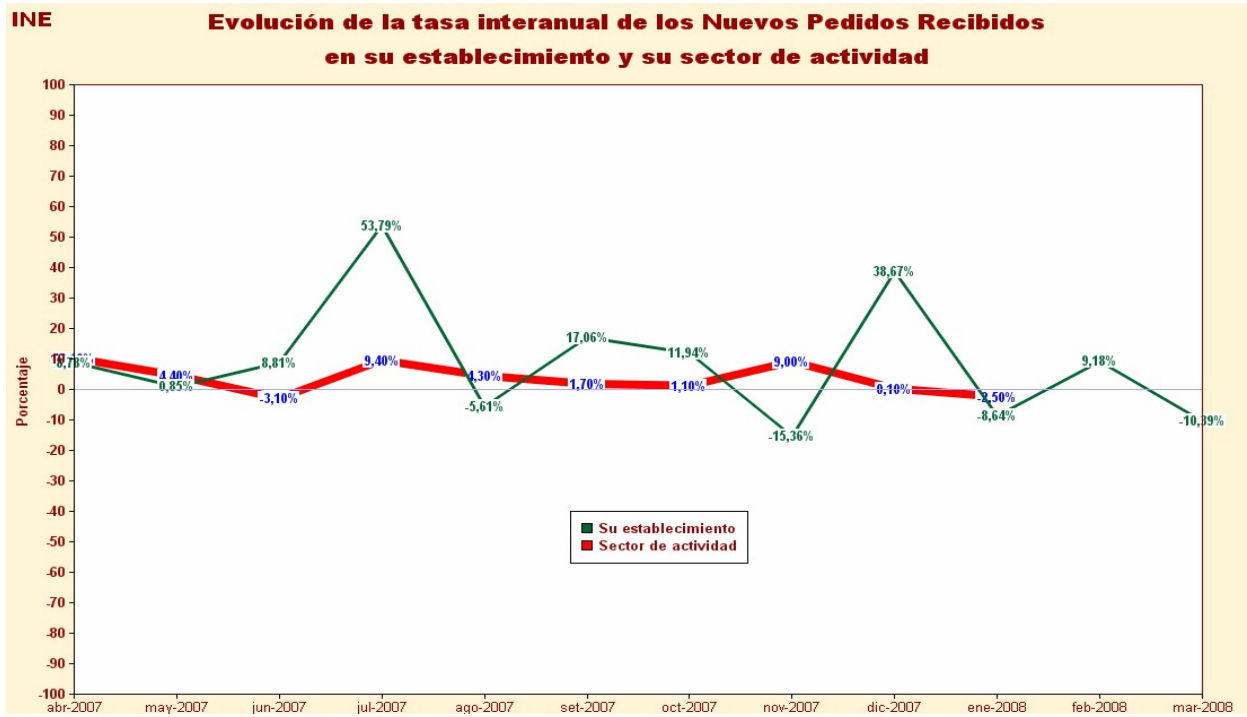


Figure 1

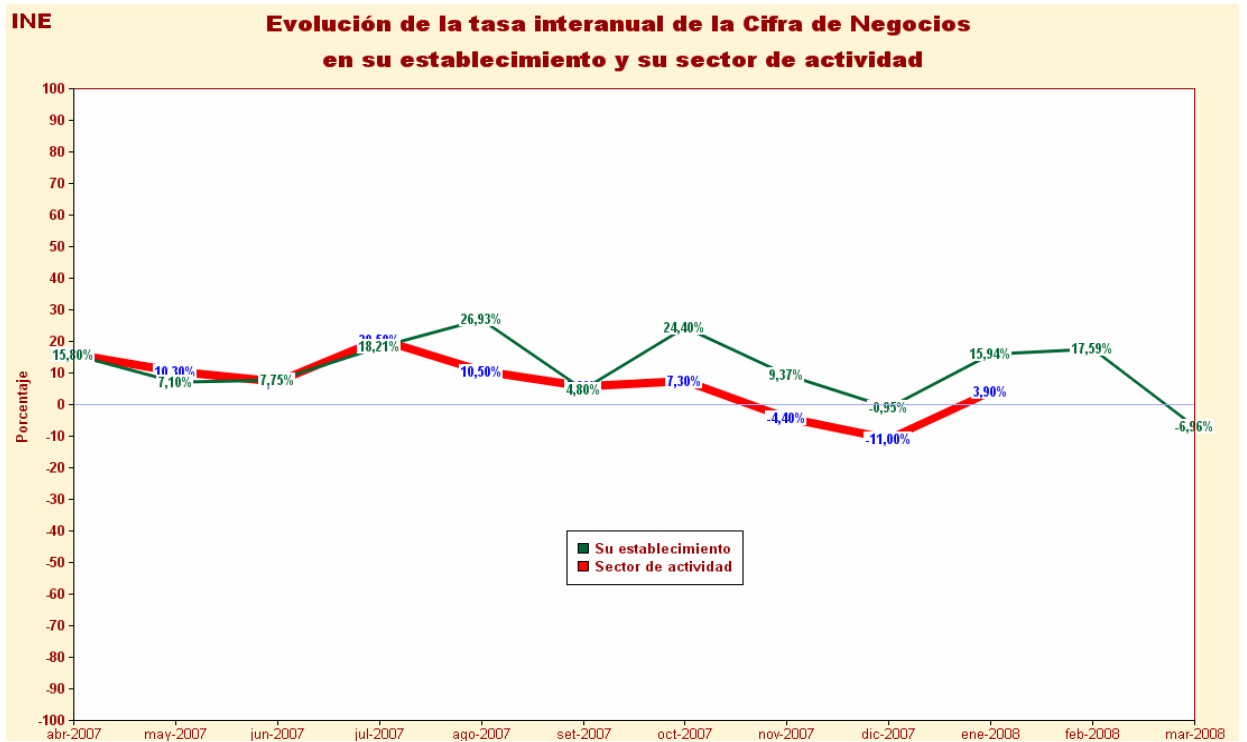


Figure 2