

**UNITED NATIONS STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR EUROPE  
CONFERENCE OF EUROPEAN STATISTICIANS**

**UN/ECE Work Session on Statistical Data Editing**  
(Madrid, Spain, 20-22 October 2003)

Topic (iv): Data editing by respondents and data suppliers

**DATA EDITING BY REPORTING ENTERPRISES**

**Invited Paper**

Submitted by the National Statistical Institute, Spain<sup>1</sup>

**Abstract**

This paper discusses the role of reporting enterprises in data editing. The approach followed is that of TQM. Hence, suppliers and customers are part of the productive system and a global strategy is needed for an efficient data editing. Some prerequisites to get high quality incoming data are discussed. The paper presents two aspects where reporting enterprises may play an active role in data editing: Web surveys and an “auditing system” carried out by the reporting enterprises. The experiences in Spanish industrial surveys are also presented.

Keywords: EDR, Web surveys, TQM

**I. INTRODUCTION**

1. Nowadays, public statistical institutes are under continuous pressure from society, which demands more and more data, to be produced at a lower cost and with a lower respondent burden. In this situation, methods that could help to improve the efficiency of statistical production processes, and to fit better customer needs, while reducing respondent burden at the same time, are of the greatest importance. Total Quality Management (TQM) is being used by many organisations all over the world, as a method to improve production processes and to produce high quality products.

2. Moving editing closer to respondent can significantly contribute to improving editing effectiveness. We can go a step further by integrating the respondents in editing processes. Using one of the main TQM principles (i.e. the suppliers and the customers are part of the productive system), we should integrate our suppliers (the respondents) in our productive systems. This paper discusses the role of reporting enterprises in data editing. Web surveys offer new possibilities on moving editing closer to respondent. The “auditing system” carried out by reporting enterprises comes from the TQM action of offering to the enterprises tailored data on their market shares as a “payment” for filling in questionnaires.

3. In the following section, some prerequisites to get high quality incoming data are introduced. In section III, a new paradigm to quality evaluation, auditing by reporting enterprises, is presented. The role of reporting enterprises in Web surveys is discussed in section IV. The paper ends with some final remarks.

---

<sup>1</sup> Prepared by Pedro Revilla (previlla@ine.es), Manuel Gonzalez (mgdavila@ine.es), Margarita Gonzalez (mgonzalez@ine.es) and José Quesada (pquesada@ine.es).

## II. PREREQUISITES TO GET HIGH QUALITY INCOMING DATA

4. It is clear that high quality macrodata cannot be produced from poor, erroneous or biased microdata. Hence, the first requisite to produce high quality statistics is that reporting enterprises transmit high quality data to the statistical office. There are no data editing method that would be able to amend low quality raw data.

5. The problem of how to get high quality incoming data can be faced from the perspective of the TQM. Using a TQM approach, it is considered that the suppliers (i.e. the reporting enterprises) are part of the productive system. Hence, the statistical process begins with the production and transmission of microdata by the reporting enterprises. Data collection has to be adapted to the respondent conditions and possibilities. Also, statistical agencies should implement corporate strategies to encourage the respondent to fill in questionnaires with confidence and care. Moreover, data editing should improve the response quality (Granquist, 1997).

6. In the INE, we have adopted a TQM approach to produce industrial statistics. Adapting questionnaires to the accounting practices of the enterprises and improving our relationship with enterprises are two key success factors in achieving high quality incoming data. We have taken the following measures in order to adapt questionnaires to the accounting practices of the reporting enterprises:

- The requested variables and the valuation rules are adapted to those of the enterprises.
- The observation units are adapted to those units for which enterprises have available information.
- Different models of questionnaires adapted to the branch and the size of the enterprises are used.
- Questionnaires are personalised.
- The transmission formats (paper, fax, diskette, e-mail) are adapted to the choice of the enterprises.

7. The underlying principle in this approach is that enterprises provide data in the same way they produce them for their own use, and the statistical agency re-elaborates them for analytical purposes, if necessary. Adapting questionnaires to the accounting practices makes answering them easier and quicker. It also results in fewer errors made by the enterprises. All of this leads to an improvement in the quality of incoming data.

8. Another key success factor in achieving high quality incoming data is improving our relationship with reporting enterprises. One of the ways this is being achieved is by offering the enterprises free of charge data tailored to their needs. This new practice (described with more detail in the following pages) has seen an increase of interest on the part of the enterprises, which are consequently filling in questionnaires sooner and with more care.

## III. AUDITING BY REPORTING ENTERPRISES

9. Nowadays, there are a lot of discussions and a huge number of papers about statistical quality, quality evaluation, quality reports, etc. Statistical auditing ( de Vries and van Brakel, 1999) has also been proposed, to check how quality management in statistical departments is functioning and how the quality of statistical products and procedures may be improved. In this context, reporting enterprises may have an active role on editing and quality evaluation. An external evaluation on the data quality, carry out by reporting enterprises, may be achieved, because of the prior knowledge of the enterprises about their sectors and markets. This external evaluation may be considered as a consequence of one of the main TQM principles: the suppliers (the reporting enterprises) and the customers (the data users) are part of the productive system.

10. In the INE, we have established an “auditing system”, carried out by 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 questionnaires.

11. 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 questionnaires (Gonzalez and Revilla, 2002). The frequent complain of enterprises on statistical burden was the origin of our offer. Enterprises often wonder what use questionnaires are to them. This dissatisfaction on the part of the enterprises (our suppliers) contrasts sharply with our customer general satisfaction. The solution that we found to this contradiction was to change our suppliers into customers! We had to look for the kind of statistical data for which reporting enterprises are more interested. We have found that usually they are not so much interested in general statistical data, but in very specific data, aimed to help them to be located in the market, related to its competitors.

12. Hence, we offer tailored data on market share, in exchange for the questionnaires. This data are computed based on the data of each enterprise, comparing it 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?

13. The reaction of enterprises to this offer has been quite positive. As an example, the president of the Industry Commission of the Spanish Confederation of Trade Association has said the followings words: “The industrial surveys carried out by the National Statistical Institute are the best example of a joint venture between industrial enterprises and public statistical offices. In other words, they are an excellent illustration of the advantages that may be achieved by means of an efficient collaboration between public and private sector.”

14. Hence, we are trying to use a new model of relationship with enterprises, that we have named, in the words of the president of the trade association, 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. The joint venture model can be seen as a particular case of the virtuous circle pointed out by Felegui (1991): “public confidence is essential to win co-operation for our surveys; this is a prerequisite for high quality statistics; high quality must ultimately be the foundation of public confidence”. Because of the increasing interest of enterprises, they are filling in questionnaires sooner and with more care. Hence, we can produce better statistics. An important point of using the joint venture model is that both timeliness and accuracy can be improved, reducing the perceived enterprise burden at the same time.

16. 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. Hence, we tried to encourage that feedback as much as possible. For example, we introduced in the sheet that we sent to the enterprise (figure 1), the following: “If this data do not fulfil your expectations, please do not hesitate to contact us. It allows our Institute to analyse them and improve our estimations. Thank you.”

17. Last year, 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.

#### IV. THE ROLE OF WEB SURVEYS IN DATA EDITING

18. Web surveys offer new opportunities on moving editing closer to respondents. Whereas Computer Assisted Interviewing (CAI) integrates into one step previously distinct phases such as interviewing, data capture and editing, Web surveys go a step further by shifting such activities to the respondent. Hence, Web surveys offer the opportunity for re-engineering editing processes, in a way reporting enterprises may play a more active role in data editing.

19. Many statistical offices are experimenting with the use of different electronic data reporting (EDR) options in data collection. Web surveys 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 (and increasingly, also in households). 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.

20. Several advantages could be expected from using Web surveys. These include improving accuracy and timeliness, and reducing survey cost and enterprise burden. 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. Moreover, this elimination of data keying reduces the processing time of the survey. There are other factors that can also contribute to improve timeliness. Data transfer on the Web can be done much faster than using the postal system. Some electronic devices (automatic data fills and calculations, automatic skipping of no applicable questions, etc.) could help the respondent to fill in the questionnaire faster. The cost for statistical offices to carry out a survey using the Web 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.

21. Nevertheless, to get the target of reducing enterprise burden using Web surveys is not so straightforward. The reduction in the enterprise burden is not always obvious. The respondents' benefits depend largely on the way metadata support the respondent in filling in the questionnaire (help texts, auto-fill rules, pre-filled data, etc). In any case, the respondents' benefits need to be clearly explained to convince them to use the Web questionnaire. An important element to improve the acceptance of Web surveys 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. 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, etc.).

22. There are a lot of expectations about the role of Web surveys in the years to come. Nevertheless, the implementation of Web surveys and others EDR methods in enterprise surveys (and, even more, in household surveys) has often been lower than expected. The take-up of electronic data reporting for statistical data by business providers is generally less than 10%, and often less than 5% (Branson 2002). Other studies also find low rates of response via Internet. For example, Gradjean (2002) finds a rate of 18% for a survey used to construct the Index of Industrial Production in France. Different rates are found in this study by enterprise size (large enterprises higher than small and medium) and by sectors (for example, electronic and electric industries more than the average, furniture industries less than the average). In another study, Mayda (2002) finds a rate between 5% and 25% in two quarterly surveys on business and agriculture in Canada.

23. More research is needed to look for the reasons why, up to now, the rate of using EDR is quite low, while technical conditions are available for many of the respondents. Probably, the electronic forms have not the same advantages for the reporting enterprises than for the statistical offices. For many of the

questionnaires, the tasks that take the most time are looking for the required data and computing their answers. There is no time difference keying data on a screen to fill in a questionnaire on paper. The advantages for the reporting enterprises could be bigger if the information can be extracted straight from their files. But this procedure may be expensive for both reporting enterprises and statistical agencies, because an initial investment is needed.

24. In any case, for most of the surveys, it is clear that, at the moment, EDR cannot be the only way of data collection. Paper data collection and associated procedures (like scanning) are 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 when using paper to an electronic questionnaire.

25. There are two contradictory targets. On one hand, to implement a single point of entry for all agency surveys, with a uniform security model and a common look across the entire site. And, on the other hand, to allow decentralised applications to cope surveys singularities. One aspect where the difference among surveys has to be taken into account is data editing. Combining the two targets (i.e. integrating a centralised platform with decentralised applications) is a non-trivial task.

26. Some crucial questions arise: What kind of edits should be implemented on the Web? How many? Only fatal edits or fatal edits and query edits? What kind of edits should be mandatory?

27. On one hand, we need to include some edits. If we do not, then the information collected by a Web survey 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 Web surveys: no need to editing again the information with a suitable set of edits implemented in the Web application. On the other hand, we need to be extremely careful in the set of edits to be implemented in the Web survey, 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 Web surveys, as users will prefer the easy way (paper).

28. How to cope with the too few/too many edits dilemma? If we are trying to implement a Web questionnaire in an existing survey, a way is to analyse the current set of edits in order to determine the efficient set of edits to be used in the Web implementation. Hence, the implementation of new procedures obliges to the revision and redesign of the current procedures of the survey. But we should make that revision *from the user's point of view*. Otherwise it is impossible to find out if the users of a Web survey (the respondents) are going to get fed up with the task of filling in a Web form or not. It must be stressed that making that sort of analysis is strictly necessary in order to implement a suitable set of edits that will not discourage users and that will make possible not to edit the Web information in the traditional paper way.

29. In order to achieve this target an analysis similar to that of Martin and Poirier (2002) should be carried out. It is important to dispose of procedures that allow having access to versions of data and additional processing metadata that describe how the data were transformed from collection to dissemination.

30. Like many others statistical agencies, the INE has a significant interest in Web-based data reporting. An example of this was the possibility offered to all citizens to fill in the Population Census 2001 using the Internet. The INE is working in a general Project of giving respondents the option of submitting their responses to statistical surveys using the Internet. 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.

31. Moreover, an ad hoc prototype Web system to collect establishment data for the Turnover and New Orders Survey is also being implemented. This monthly survey uses a very simple form that includes only two variables: turnover and new orders, broken down by geographic markets (domestic, EU Euro-zone, EU non-Euro-zone, and non-EU). The following data are also collected on orders: stock of

orders at the beginning of the reference month, new orders, cancellations, orders invoiced and stock of orders at the end of the reference month.

32. Many problems using the Internet might be due to the various configurations and products installed on the respondents' machine. Each respondent's computer can have different components, different versions of operating systems and browsers, and different modem speeds. For this reason, and from using a very simple form, all the programs are going to be run through the server, without the need to install any software on the respondents' computer.

33. We think that very few edits should be mandatory. Hence, in the majority of the cases, the respondent has the option to ignore the edit message and submit their data. The only mandatory edits are: at least some data should have been entered in the questionnaire, alphabetic characters are not allowed in numeric fields, and three accounting equalities (i.e., fatal edits). These include the following: total turnover (new orders) equals the sum of turnover (new orders) by the different markets, and stock of orders at the end of the month equals stock of orders at the beginning of the month, plus new orders, minus cancellations, and minus orders invoiced.

34. The rest of the edits are included as a message, which is sent to the respondent. These include "some fields are missing" and historic edits (monthly and annual rates). A fatal edit, "stock of orders at the end of the month  $t-1$  equals orders at the beginning of the month  $t$ ", is also included, only as a message, and not as a mandatory edit. The acceptance intervals for historic edits are wider than in traditional microediting, trying to minimise the number of messages.

35. The idea for the future is that, after Web edits, no traditional microediting is needed. A selected editing approach based on time series modelling (Revilla 2002) will be 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.

36. The Web form will be offered to the sample of reporting enterprises as a voluntary option to respond to the Survey. We think that, probably, many enterprises will not change to the Web form. For this reason, we offer tailored data to them, following the approach presented in the previous section. When an enterprise sends a valid form (i.e. passing the mandatory edits), it immediately receives tailored data from the server. These tailored data consist of tables and graphs showing the enterprise trend and its position in relation with its sector. Offering this data through the Web has some advantages (speed, possibility to edit the file) over sending this same data on paper by mail. Some examples of these data can be seen in Figures 2 and 3. Taking into account these advantages, we expect more enterprises to use the Web survey.

## V. FINAL REMARKS

37. Probably, we are still far from the desired "joint venture model", where reporting enterprises may play a more active role in editing and quality evaluation. Also, the number of electronic questionnaire users is too small at the present time. Moreover, more research is needed on selective editing methods (for example, the use of statistical modelling to construct more efficient edits). But the combination of the TQM approach, EDR methods and selective editing strategy, may be in the near future a key success factor in producing high quality data, at a lower cost and with a lower respondent burden.

## REFERENCES

Branson, M. (2002), "Using XBRL for data reporting". Australian Bureau of Statistics. Unece/Eurostat Work Session on Electronic Data Reporting, Working Paper No 20. February 2002.

De Vries, W and Van Brakel, R. (1999). "*Quality systems and statistical auditing: a pragmatic approach to statistical quality management.*" Quality Work and Quality Assurance within Statistics. DGINS Conference in Stockholm.

Fellegui, I.P.(1991) "*Maintaining Public confidence in Official Statistics*". Journal of the Royal Statistical Society. A.154. Part1. pp.1-22.

González M and Revilla P., (1997). "*Total Quality Management and the INE*". Eurostat Web.[www.forum.europa.eu.int/](http://www.forum.europa.eu.int/)

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

Gradjean, J.P. (2002), "*Electronic data collection in Official Statistics in France*". French National Institute of Statistics INSEE. UN/ECE/Eurostat. Work Session on Electronic Data Reporting. Working Paper No 7. February 2002.

Granquist, L. (1995). "*Improving the Traditional Editing Process*", Business Survey Methods. Wiley Series in Probability and Mathematical Statistics.


Granquist, L. (1997) "*On the Current Best Methods document: edit efficiently*". UN/ECE Work Session on Statistical Data Editing, Working Paper No 30. Prague. October 1997.

Martin, C. and Poirier, C. (2002), "*Analysis of data slices and metadata to improve survey processing*". Statistics Canada.. UN/ECE Work Session on Statistical Data Editing. Working Paper No 2. Helsinki. May 2002

Mayda J.(2002), "*Experiences with implementation of EDR into existing survey programs 2002*".Statistics Canada. UN/ECE/Eurostat. Work Session on Electronic Data Reporting. Working Paper 23. February 2002.

Revilla P.(2002), "*An E & Method based on time series modelling designed to improve timeliness*". National Statistical Institute. Spain. UN/ECE Work Session on Statistical Data Editing. Working Paper No 13. May 2002.

INSTITUTO NACIONAL DE ESTADISTICA



ENTERPRISE XXXXXXXX

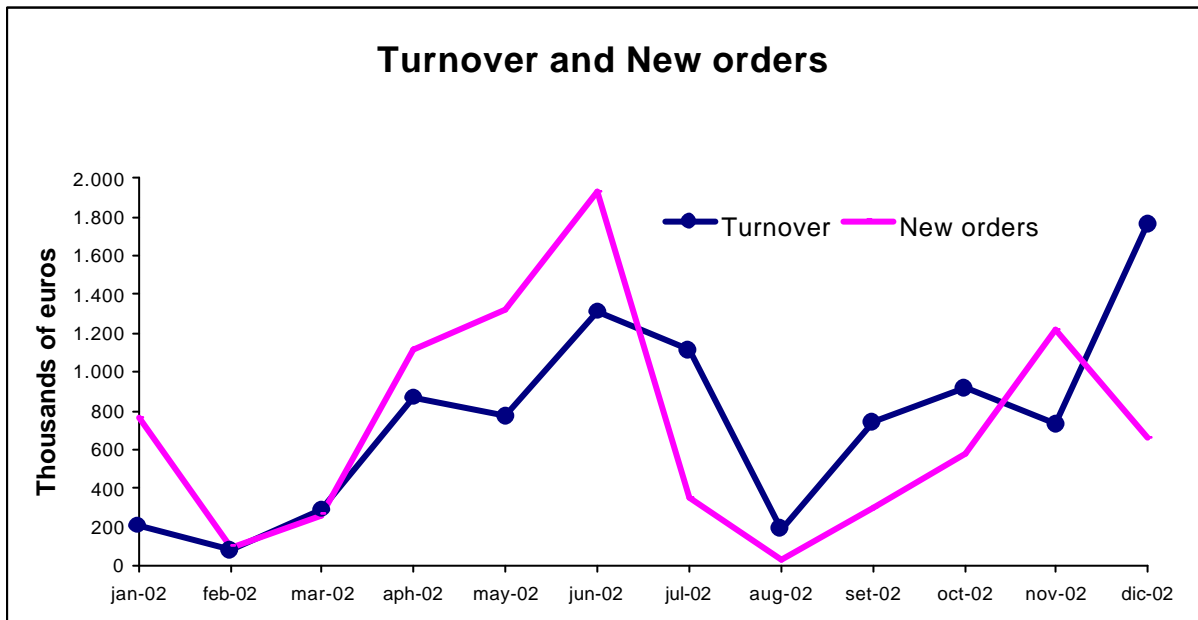
BRANCH : 15.33 - Processing of fruit and vegetables

Number of enterprises	895
Enterprise market share	0.25 %
Number of enterprises with larger market share	101
Market share of that enterprises	68.41 %

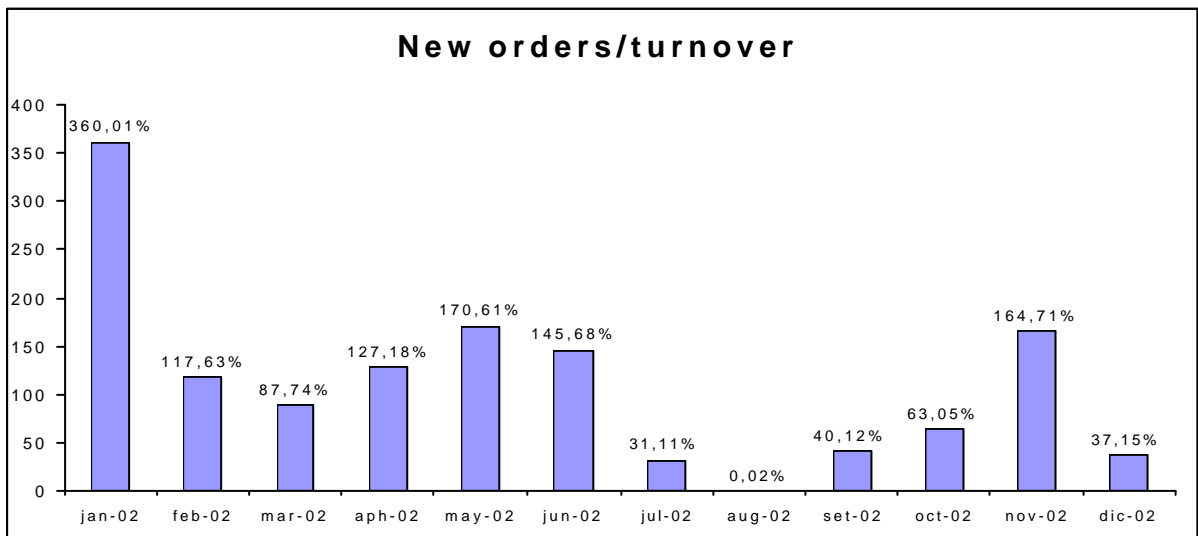
This data is for your information only.

If this data does not, fulfill your expectations, please do not hesitate to contact us. It allows our Institute to analyses them and improve our estimations. Thank you.

**Figure 1**



**Figure 2**



**Figure 3**