

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Prague, Czech Republic, 14-17 October 1997)

Item 5 of the provisional agenda

CHANGING THE SURVEY DESIGN AFTER IMPLEMENTATION

Submitted by the Statistical Office of the Republic of Slovenia

¹ Prepared by Pavle Kozjek.

I. INTRODUCTION

1. The recent development of statistical methods and techniques has opened many new possibilities in survey processing. Modern tools were developed, on the one hand, to help statisticians and informaticians to get results - statistical information - in shorter time and with less effort. But, on the other hand, less time, money and human resources are available to get the same statistical information. Timeliness became a problem for many developers, and it usually results in a lack of time to completely design and test the application. With CAI (particularly CAPI) applications this can lead to serious problems when such an application is installed to many laptops, and some data are already collected and edited. Changing data definition means incompatibility of data files, which usually cannot be solved without developers.

2. This paper describes one of the possible solutions to this problem when the survey application is already in production. The reported application is based on Blaise III software. At the Statistical Office of Slovenia (SORS) it was successfully used in the continuous Household Budget Survey (HBS), conducted since February 1997.

II. SLOVENIAN HOUSEHOLD BUDGET SURVEY - a short description

3. HBS has been carried out in Slovenia since 1963 on a larger sample (every five years about 3200 households) and on a smaller sample (every year about 1000 households). Since 1997 the Slovenian Statistical Office started with a continuous HBS. Every year about 1200 (net) households will be interviewed. Since this sample size is too small for more detailed calculations, the data will be aggregated together for 3 years. Every year the "oldest" 1200 households will be eliminated from the sample and the "newest" 1200 households will be included.

4. The survey design was radically changed, making it more regular, with a new questionnaire, new number of interviews, and diaries were also introduced. Data processing was completely redesigned. A pilot survey using Blaise III was conducted in October 1996. In February 1997 the preparation of an HBS was launched.

5. In comparison with the annual Household Budget Survey, the continuous HBS was expected to have some advantages:

- higher quality of current data and lower costs;
- better organization of the field work and methodology of the survey;
- better trained interviewers;
- more detailed results, derived from more subsequent years aggregated data.

6. The questionnaire is divided in two parts and interviewers make two visits to each household. At the first visit, the first part of the questionnaire is completed and the diaries are explained. After two weeks the second part of the questionnaire is filled in and the completed diaries are collected.

III. THE SURVEY APPLICATION DESIGN

7. The main problem with the Household Budget Survey was the lack of time for application development. HBS is one of the most complex surveys: the paper questionnaire consists of 72 pages and over 3000 possible fields to be entered. The final version (February 1997) of the CAPI-Blaise application consisted of 98 screens, mainly tables. But there were only about four weeks and one person available to develop the Blaise instrument, together with the entire case management.

8. The processing of the HBS questionnaire is completely independent from the processing of diaries. The questionnaire is processed in Blaise III as a multi-mode survey: there are 22 interviewers with laptops, using CAPI instrument and 4 data entrists, entering data from paper forms (PAPI/CADI). After coding of PAPI/CADI data (CAPI data are coded by interviewers), all data files are combined into one file and sent to the SAS system for further processing. CADI application for diary processing is still in Blaise 2.5. Coding is the most important part of it. All three available types of coding are enabled: alphabetical, hierarchical and trigram.

9. In order to respect the deadlines and to train interviewers, the application had to be installed on laptops on the same day when the last question was coded into the Blaise language. There was no time available for testing. Furthermore, due to the lack of time, only the most important edits were included in the first production version of Blaise III.

10. Small errors were corrected already on the first day of the interviewer's training. The interviewers were expected to start data collection and editing immediately after the second training session. After the second installation, the absence of some important edits, which should be performed during the interview, was identified. Interviewers could start their job in time, but it was necessary to correct the data entry and editing application. After a month running, the application was tested satisfactorily.

IV. RE-INSTALLING THE SURVEY APPLICATION

11. When re-installing the HBS application the instructions from the Blaise III Developer's Guide were followed. Owing to some difficulties in newly established modem communication between interviewers and the statistical office, it was decided to send a new installation diskette to each interviewer, together with short printed instructions.

12. Because of the complexity of the questionnaire, data conversion was feared, and especially the burden placed on interviewers. With this in mind, the re-installation procedure was prepared very carefully and made as simple as possible - a real "push-one-button" solution. Interviewers were only asked to start the batch procedures which installed some files and new main interface menu, where they had to choose the new option called "Conversion to a new data definition". This option executed the following steps:

- Creating the OLD directory and copying the existing data files into it;

- Installing the new improved survey application to a working directory;
- Installing and executing of the Manipula conversion setup.

13. The Manipula setup to convert data from an old to a new data definition can be generated automatically any time during survey processing on the basis of meta-data definitions. Therefore, the conversion procedure can be executed more than just once, if necessary. Automatic generation of Manipula setup is the key to the entire procedure. The development of the conversion application would be much too time consuming. The interviewers initiated the new application in the same way as the old one. A backup version of old data and metadata was automatically saved in a separate directory.

14. Conversion to the new data model was quite successful. It was possible to execute the conversion without losing the necessary files. The same procedure was performed on a centralized part of the survey application on PAPI/CADI data files at the statistical office. Old and new data definitions were the same as in the CAPI instrument, and new data files from both modes of data collection were combined into one file again. The new SAS setup was produced from the new data definition and data were sent to further processing.

V. INFLUENCE ON NEW POSSIBILITIES TO THE EDIT STRATEGY

15. A unified strategy for an improved editing process starts at the design phase and ends only when the final results are published. The aspects of quality of statistical information are accuracy, timeliness and lowest costs. The new survey design could improve especially timeliness. Which benefits can be obtained from using this possibility? Edits can be added or removed during the survey data processing. Development and production of the survey can run concurrently.

16. There are, however, also some disadvantages such as, for example: additional interviewer's burden, adding new fields usually means repeated interviews, and more survey administration is needed.

17. As for the HBS, the advantages were clearly prevailing, so this or a similar approach will probably be used also in some other surveys at SORS, especially when rapid application development is necessary.

VI. CONCLUSION

18. Growing needs for faster and more accurate statistical information are often in conflict with the developer's possibility to complete and test the survey application in time. With tools and methods which enable completing and finalizing the survey even after its implementation, some extra time could be acquired. In many cases, only a small amount of extra time means the difference between the interviewer's exhausting hard work and easy, user-friendly data collection and editing. The possibility of redefining the survey after implementation can improve timeliness as well as the quality of collected data.

References

Bethlehem, J. (1995), "A Control Centre for Computer-assisted Survey Processing". Report, Voorburg: Statistics Netherlands.

De Jong, W.A.M. (1996), "Designing a Complete Edit Strategy; Combining Techniques". Research paper no. 9639, Voorburg, Statistics Netherlands.

Schou, R. (1995), "Developing a Multi-Mode Survey System". Third International Blaise Users Conference, Helsinki.