CAPI and the Collection of Living Conditions Data; A User's View

Ari Tyrkkö and Hilkka Vihavainen, Statistics Finland

1. Introduction

The Living Conditions Survey 1994 was the first comprehensive survey, where the CAPI-method was widely applied in Finland. The sample and data collection of the survey were combined with the Income Distribution Survey.

The decision to start with the new method and to combine two different types of surveys was taken one year before the beginning of data collection. The computerisation of the field interviewers was considered such a big investment that it needed the backing of a real data collection project, not merely a pilot. At that time, the Income Distribution Survey was the only annual survey with personal interviewing. The project was ambitious, because the CAPI-organisation had to be created from the very start (see [1]). Furthermore, the period for building up a totally new data collection system was very short.

The field work was preceded by a pilot survey with a sample of 1 000 individuals. Half of the persons were interviewed by the CAPI-method and half with the conventional PAPI-method. The aim of this work was to collect comparable data with both methods for methodological research in order to study the effect of the interviewing method on living conditions data. The results of this study will be reported in the near future.

In general there were no technical problems with the use of the CAPI-method (with BLAISE) in the individual-based Living Conditions Survey. The difficulties, which came up with the new method, were mainly related to the organisation of the survey and the documentation of data collection instruments. The Living Conditions Survey is conducted in Finland every seventh or eighth year. The preceding one was from the year 1986. Less than half of the questions were maintained in their old form.

The household-based Income Distribution Survey had a more complicated questionnaire structure containing questions addressed both to the whole household and its individual members. The problems which arose in this connection were different in nature. The Income Distribution Survey is conducted in Finland on an annual basis. More than 90 percent of the income items are gathered from different registers. The main purpose of the interview is to collect information on household composition, on the economic activity of the household members and on some rare income items, which are not registered. The Income Distribution Survey is based on a two year panel.

2. Programming of a BLAISE-questionnaire

The programming of an individual-based questionnaire was quite straight forward. The BLAISE-method did not set any limitations to the structuring of the questions, which were presented earlier in the previous PAPI-interviews. Only a part of the more complicated question groups had to be remodified.

The programming of household-based questions in the Income Distribution Survey was a much more complicated task. The old PAPI-based formulation of some questions was impossible to translate into BLAISE-form. This concerned especially the rosters of household members and their economic activity.

Though the BLAISE is developed partly as a tool for a researcher with no ADP-background, the programming of a complicated or comprehensive BLAISE-questionnaire could be carried out only by an ADP-expert. An ADP-expert was used in the programming of both parts of the survey (Living Conditions and Income Distribution). The reasons for this decision were mainly practical caused by the short time period for the planning of the survey. Anyway, working with the new tool seemed much like the learning-by-doing method - as usual.

3. Data processing

In the earlier PAPI-based production systems interviewers sent all the questionnaires to the central unit, where the data was coded manually and checked centrally with the computer. The clean data was combined with data from registers and other data sources. The production took place in an IBM-mainframe system.

Part of the earlier centralised data checking procedures were transferred to BLAISE-questionnaire to be conducted during the interview. As a consequence the data was obtained 'cleaner' than before, when it arrived from interviewers.

In earlier PAPI-interviews the interviewers made their remarks on the margins of questionnaires. These remarks were checked afterwards separately household by household during the data processing. In CAPI-interviews the respective remarks were gathered under certain code numbers. The checking of these remarks was more effective, because it could be conducted question by question. The remarks also gave easy-to-access information on the quality of questions.

The coding of some classifications was transferred to the CAPI-questionnaire. This concerned the computer-assisted coding of municipalities and occupations. Especially the coding of occupations was supposed to give better results during the interview than afterwards centrally. The system was built up on the dictionary coding-program of BLAISE 2.4. However, it proved out to be too mechanical and involve still a lot of development: The Scandinavian letters were missing and the right items of occupation were difficult to find. This time all the classification numbers had to be checked afterwards manually.

The data transmission (from interviewers to the central unit) system turned out to be problematic for the data processing (see [1]). Whenever interviewers sent new data to the central unit, their whole data file was transmitted including all the earlier concluded interviews. This system enabled the interviewer to make corrections to the old interviews almost freely. It also meant that the data was not ready for processing until the last interviewer contact was made. All in all, this system had the result that most of the advantages of the more effective data collection method for time tables were lost. Earlier when the data was collected by PAPI-questionnaires, it was processed immediately after the central unit had received the questionnaires. Now they had to wait for the last questionnaire. As a result the entire production timetable was delayed by two months from the planned schedule. However, we were able to publish some results of the Living Conditions Survey about half a year earlier than in the previous survey.

One of the subjects, which still needs a lot of development work in the production of the Income Distribution Survey, is how to transfer a large amount of data from BLAISE into the Datacom database, which is in the IBM-mainframe.

The sequential files that were produced with BLAISE-program were rigid to use for further analysis. This concerned especially the Income Distribution Survey. The sequential file that was produced from the questionnaire was 8 000 bytes long, most of which was empty space. This was caused by the household - person roster of the questionnaire. For every potential member a field had to be reserved, if the question was an individual-based one. It meant a space for 16 potential persons per question.

Panel surveys contain the possibility of using the information, which has been collected earlier, in interview situations. This characteristic would have been one of the advantages of the CAPI-method. However, this time it was not utilised.

4. Survey organisation

The introduction of new methods into on-going production systems within a short period of time sets high requirements for the co-operation of people. One of the consequences of the change from PAPI to CAPI/BLAISE was a new division of labour between ADP- and statistical experts. In our case the programming of the questionnaire was left mainly to people specialised in ADP. Sometimes the process appeared to be like a 'black box' to the statistical experts. The functioning of the questionnaire in different situations was difficult to test. If the programming of the questionnaire would have been transferred solely to the statistical experts, it would have reduced the share of ADP-experts' work in the production system and lead to another kind of division of labour.

One of the problems with the survey organisation was the documentation of the questionnaires. A paper copy of a more complicated CAPI-questionnaire with modifications according to different household structures and answers to certain questions is impossible (in practice) to produce. If a survey is conducted regularly but with longer intervals (e.g. every fifth year) the documentation has to be careful. The interviewers as well as some research oriented data users need also some kind of flow chart of the contents of the questionnaire or a list of questions in order to get a total view of the questionnaire. Due to the comparative

survey of interview methods a PAPI-questionnaire of the Living Conditions Survey was produced (but only for this once). The questions from the Income Distribution Survey were also listed on paper.

5. Conclusions

There were three main goals in the implementation of the CAPI-method in the Living Conditions Survey: to collect the data in conjunction with the Income Distribution Survey; to meet the data quality requirements; and to speed up the data collection process.

The first goal was achieved although the design of the survey was very complicated and gave rise to quite a lot of discussion and work. The quality profile of the data has proved to be fairly good. The method did not have a considerable impact on non-response and on major distributions, and the quality of the data has presumably improved. The third goal was met only partly. Although we expected to get the data ready earlier than we actually did, we can feel satisfied with the speed the results became available.

Still, two important lessons remain to be learned. The integration of two large scale surveys was a very demanding task - maybe a bit too demanding. The linking of two different types of questionnaires to one interview session led to a comprehensive and complicated instrument, which was difficult to control. Another main difficulty was the documentation of the questionnaires. Since the research oriented users of the Living Conditions Survey data are interested to know the design of the questions and the structure of the questionnaire, a paper version is necessary. A paper copy from a CAPI-questionnaire is, if not impossible to produce, at least impossible to understand for a non-expert.

Much more emphasis should have been put on the planning of the project. For the sake of the short planning period, the development of new methods was concentrated on the data collection phase of the survey. Yet the full benefits of the new methods could only have been reached, if the whole production system had been trimmed. This concerns especially the speeding up of timetables. The data, which the interviewers send to the central unit, is immediately ready for further analysis and for the publication of preliminary results, at least in principle. The 'on-line' system is possible only when the production system has been prepared for this purpose beforehand.

Reference

[1] Kuusela Vesa (1995): *Interviewer Interface of the CAPI system of Statistics Finland.* In this book.