

Experiences of usage of Blaise in CATI surveys

Stefania Macchia & Manuela Murgia (ISTAT, Italy)

1. Introduction

The paper shortly describes the positive results obtained through the use of Blaise in Istat for CATI surveys, which lead to the definition of a “standard” strategy that has already been largely adopted for three years. This strategy is based on the use of Blaise and on its interaction with other software packages, in order to provide solutions for those requirements common to all CATI surveys. Both standardised and not standardised solutions implemented in the software procedure will be described, although the last ones would more efficiently meet the CATI surveys’ needs if they were directly provided by Blaise.

2. The use of Blaise in Istat for CATI surveys

A lot of surveys in Istat are carried out with CATI technique and for a great number of them a strategy has been defined aimed at improving the quality of data. This strategy consists in relying on private companies only for the call centre and the selection of interviewers and in providing them with the entire software procedure to be used for the data capturing phase, which is based on the Blaise system.

The majority of surveys carried out with this strategy regards households and individuals, but this is only due to the fact that they are often characterised by very complex questionnaires in terms of number of questions and consistency rules, so there is a greater interest in developing in house the software procedure.

As it can be seen in the following table, a survey on enterprises (*Water system survey*) has been treated too, which has a very huge number of variables (30,000), even if this number is only a maximum value defined considering the maximum numbers of different entities which could belong to each respondent/form (any respondent/form reaches this maximum values for all the entities).

Table 1- Surveys carried out with the strategy based on Blaise

Surveys		Number of variables of the electronic questionnaire	Number of interviews
	Long	677	16,597
<i>Sample births survey</i>	Short	170	33,838
<i>Year 2001</i>			
<i>Sample births survey</i>	Long	707	15,642
<i>Year 2004</i>	Short	199	33,515
<i>University-to-work transition survey and perspectives</i>		218	25,510
<i>Year 2004</i>			
<i>Upper secondary school graduates survey</i>		315	20,408
<i>Year 2004</i>			
<i>Water System Surveys (preliminary survey)</i>		30,000	1,320
<i>Year 2006</i>			
<i>Violence against women survey (in progress)</i>		2,774	25,000 (to be concluded)

This strategy turned out to be successful as it can be deduced by different aspects which have been evaluated. One of these is the response rate, which is surely high, considering the length of the interviews (table 2).

Table 2- Response rates and interviews' lengths of the surveys carried out with the strategy based on Blaise

Surveys	Interviews' length	Response rates
<i>Sample births survey</i>	<i>Long</i> 12'00"	92.6%
<i>Year 2001</i>	<i>Short</i> 5'00"	93.2%
<i>Sample births survey</i>	<i>Long</i> 13'48"	94.7%
<i>Year 2004</i>	<i>Short</i> 5'43"	96.8%
<i>University-to-work transition survey and perspectives</i>	10' 56"	95.8%
<i>Year 2004</i>		
<i>Upper secondary school graduates survey</i>	13' 20"	94.7%
<i>Year 2004</i>		
<i>Water System Surveys</i>	9'03"	99.8%
<i>(preliminary survey)</i>		
<i>Year 2006</i>		
<i>Violence against women survey</i>		
<i>(in progress)</i>	26'54"	72.4%

Having adopted this strategy with success for more than three years, made us cope with different requirements expressed by each survey that we solved either exploiting the potentialities of Blaise or integrating the software procedure with other software packages.

Some of these solutions turned out to constitute standard tools in our procedures, some others needed to be used only for particular purposes, so we did not develop standard functions, but we identified the software packages to be used for them, while for other requirements, we would like to have at disposal some specific functions included in the Blaise system. The last ones are those requirements which, even if are not common to all the surveys, are typical of a well identified subset of them and would require a big programming effort to be integrated in a software procedure.

3. Standard tools included in the used software procedure

The standard tools we implemented in the software procedure of the described strategy are those which fulfil the requirements which are common to all the CATI surveys.

They mainly concern:

- the production of a certain number of reports to monitor daily the interviewing phase both from the qualitative and the quantitative point of view,
- a standard procedure for the electronic transmission of data (interviews and reports) from the external companies to Istat, which guarantees all the security requirements,
- a function which allows the display on the screen, during all the interview, of the called telephone number and name of the person.

Considering the first item of the list, the set of reports have been developed using **Visual Basic**, based on an Access database, which produces Excel files, that can be easily managed by statisticians.

A lot of these reports contain indicators which allow to examine more in details some elements regarding the CATI contact results already provided by Blaise through some video screens displayed by the BTMana. Some of them are aimed at monitoring the productivity of interviewers, so they are calculated per interviewer (i.e. refusals, definitive interruptions, etc), others reports are aimed at detecting possible problems of the survey questionnaires (i.e. definitive interruptions per questionnaire section) and other are aimed at optimising the proceedings of the interviewing phase (i.e. future appointments per time slices, future appointments per time slices per interviewers when the “route back to interviewer” criterion has been selected).

Other reports regard information which are stored in Blaise, but not displayed and are aimed at the already mentioned scopes. The most important are listed:

- refusal causes and definitive interruptions causes (total and per interviewer)
- interview’s length (total and per interviewer)
- last contact results distribution per type and time slice
- last contact results distribution per type and per day of week
- still available sample units classified per stratum
- matrix of movements of interviewed units from a stratum to another

Then, there are other sets of reports, regarding indicators produced through some algorithms calculated on data stored in Blaise:

- rates in total and per interviewer (response rate, refusal rate, definitive interruption rate)
- no more available sample units (because the maximum number of calls has been reached).

Some examples of reports are shown in the Appendix.

The second standard tool included in the software procedure for CATI surveys concerns the electronic transmission of data from private companies, in charge of interviewing, to Istat. The transmission is made daily, so that the Istat survey experts can monitor the interviewing activity and immediately put in practice possible correctives if something does not work well.

Three datasets are managed by this procedure:

- an ASCII file containing the Blaise “.bdb”
- the Blaise “.bth.”
- the complete set of reports “.xls”.

From the technical point of view, each day, when the interviewing activity is over, an automatic procedure, installed on a pc of the external company, downloads the Blaise “.bdb” in ASCII and includes all the above mentioned files in a “.zip” file.

Then, the private company connects to an Istat web site and, by an user_id, accedes to the “send” function. The transmission is based on a “secure” protocol (HTTPS) and puts data on an Istat server (outside the firewall) devoted to data collection. An automatic procedure transfers data on another server inside Istat firewall, where the responsible of the survey can accede (with his proper user_id) to get data.

When data are transmitted, two e-mails are automatically sent, to the company and to the Istat survey experts, so that each of them is informed of the result of this activity.

The last tool implemented for all the CATI surveys carried out using this strategy is a function which allows to show on the PC screen of the interviewers the called telephone number and the name of the respondent, during all the interview. As a

matter of fact, Blaise automatically displays these data only in the “*dial screen*”, so that, if there is a problem during the interview (i.e. if the telephone line goes off or if the interviewer wants to take note of the telephone number for any reason), the interviewer could not read this information. Technically speaking, for this purpose, a *.dll* was created with *DELPHI*.

4. Tools implemented for requirements typical of certain surveys

Other requirements are proper of certain surveys. For this kind of subjects, as already said, we did not develop standard procedures, but we identified the software tools to be used.

For instance, for two of the surveys on individuals we carried out with the above mentioned strategy (“University-to-work transition survey and perspectives” and “Upper secondary school graduates survey”), we performed the coding of Occupation during the interview. It was a particularly delicate activity, since the Occupation was the central variable of the survey. We used the Blaise assisted coding function, building a very rich *coding dictionary* (list of descriptions with correspondent codes) and we monitored the performances of interviewers on this subject through the use of control charts developed with the SAS QC software. In more details:

- for the assisted coding phase:
 - we built a dictionary of 16,300 records, each containing the Occupation description and the associated code;
 - the coding during the interview was done using the trigram search method;
- for the monitoring phase:
 - we built an application with the QC module of the SAS software;
 - the application read the database and the history file;
 - every week it produced graphical trend - on average and for each interviewer- of three main indicators - : Quality of Coding (complete codes assigned over the total number of codes to be assigned); Time for Coding; Response rate;
 - interviewers that showed systematic out of control values for one or more indicators were re-trained and monitored till their trend was close to the average one.

The entire procedure can be easily adapted to each survey that requires the assisted coding of Occupation.

5. Non standardised solutions for recurrent requirements in many surveys and that could be provided by Blaise

In this paragraph we will describe the solutions we have adopted to meet a certain number of requirements that, although are not common to all the CATI surveys carried out in Istat with the above mentioned strategy, are recurrent in a lot them. The implementation of some of these solutions required us a big effort in terms of time and resources and the solutions we found were, generally, not immediately portable to other surveys. This big effort as well as the non portability of the results might be overcome if these requirements were met directly by functions integrated in the Blaise system.

5.1. How to assign a contact result to household’s members

In household surveys, it is very frequent to extract a sample of individuals from a sample of households. The sample of individuals is determined during the interviewing phase according to eligibility criteria. For example, in the “Labour

Force” survey for each household, the eligible members are those aged more than 14; in the “Violence against women” survey only female individuals aged from 16 to 70 can be interviewed.

In these cases, to obtain a complete interview, it is necessary to contact more than one member of the household and to assign a contact result both to those who are eligible and have completed the interview and to those who are not eligible or do not wish or cannot co-operate (example: the extracted eligible member has a serious illness or refuses to provide answers or he/she is absent for all the survey period, etc.)

To meet this requirement we have developed a function that, in the course of the interview, without abandoning the form, has to assign a contact result to the extracted member and then to select the next eligible one. The implementation of this solution was very difficult because quite often its rules were in contrast with the logic of the Blaise scheduler. This happened each time the treatment associated with the contact result of the individual was different from the one associated with the contact result of the household/form. The following example tries to clarify this point.

Example: Let us suppose the household has two eligible members. The first refuses to cooperate and the next completes the interview. The treatment that Blaise would associate to the first individual (and therefore to the form) would be of “No Answer” and, therefore, the household would not be contacted anymore; what we need instead was to make the household available for the next contact in order to get the interview.

It would be therefore quite important that the Blaise scheduler could manage two kinds of dial results: one for the household/form and one for each member of it. The solution we implemented was not only very time consuming, as previously said, but also not so easy to adapt to the next survey that will have the same requirement.

5.2. How to make the structure of the datamodel of a complex survey unit more flexible

The already mentioned structure of the household (composed by a certain number of individuals to be interviewed) implies another problem: as Blaise organises the household datamodel on a single record and the number of individuals of the households is not a constant, it is necessary to define a maximum value for this number. As a consequence of this, we have very long records with a high number of empty variables.

This problem is even greater for surveys on enterprises, when it is necessary to interview the “mother enterprise” and its local units (or function units), because the maximum number of these units of lower hierarchical level is even greater than the maximum number of individuals of a household.

A possible solution could be of defining a main datamodel corresponding to the unit of the superior hierarchical level (i.e. the household) which calls other datamodels corresponding to the lower hierarchical units (individuals), so as not only to assign contact results to each of them (see point 4.1), but also to write a single record for each of them.

We are trying to implement such a structure for the second part of the “Waters system survey”, managing the interaction between different datamodels with **Visual Basic**.

5.3. How to make the managing of parallel blocks more flexible

The great majority of the CATI surveys uses parallel blocks to assign a dial result and, to this aim, an easier Blaise management of them would have better satisfied some surveys' needs; the main of them are described in the following.

Some surveys use more types of dial results than the eight ones provided by Blaise and for this reason different dial results can be associated to same treatment. Besides, it is quite common to associate the same parallel block/treatment to a dial result that can be keyed both from the dial screen and from the datamodel. This normal situation becomes difficult to manage under these circumstances:

the parallel block has to perform different actions according to where the dial result is assigned (from the dial screen or from the datamodel);

the data model includes more than one parallel blocks associated to the same treatment.

In this situation, once a dial result is keyed from the dial screen it can potentially be linked to any of the many parallel blocks that have the same treatment.

In our experience, we have to face this for the appointment treatment that was linked to three dial results. One of them was "answering machine with message" - the message said the time or period the family was at home and the operator has to fix an appointment according to it-; this dial result was present in the dial screen. At the same time, the datamodel allowed also to take other dial results associated to the appointment treatment: "appointment with the family" and "appointment with the extracted member". Besides, the parallel blocks associated to the three cases had to perform different sets of rules.

To overcome the problem that it is not possible to link the "answering machine with message" result to its specific parallel block (because all the parallel blocks have the same treatment), we had to use only one parallel block that contained all the rules relative to all the three cases and put all these actions under a very complex system of "*if-and-then*" conditions that allowed to perform the right rules according to the selected dial result.

Probably this could be avoided if the Blaise system would have allowed to link each dial result not to a treatment but to the a customised name of the parallel block associated to the treatment.

Another very frequent requirement we have to meet is to assign different dial results according to when the phone contact is closed. For example: for all surveys, the dial result "refuse" can be selected only until the very beginning of the interview (when administering some "screening" questions), while, when questions concerning the proper survey subject start, the dial result is "definitive interruption" in case the respondent does not wish to co-operate anymore,. In the first case we link the dial result to the treatment "No Answer" while in the second case to the treatment "Other". As we do not want to make the operator choose between them (as we noticed there are frequently misunderstandings) we make possible to key the right result making appear only one of the two parallel blocks a time, depending on which part of the questionnaire is administered.

To do this we have to put under "*if-and-then*" conditions not only the two parallel blocks involved but also the entire set of parallel blocks used throughout the questionnaire. This implies that we have to find a suitable condition for the calling of each parallel block, also for those that would normally not need it, like the "No answer".

Clearly, in both the situations described, the complex system of “*if-and-then*” conditions is peculiar to each CATI survey because it is generally based on values of the survey questions. It is therefore almost impossible to create standardised set of rules in that they have to be customised to the survey questionnaire.

A parametric management of the parallel blocks in Blaise could probably represent an easier way to meet this requirements.

5.4. How to modify the Blaise database

The need of a change in the variables’ values on the Blaise database while the interviewing phase is in progress is very common to all the CATI surveys we manage.

These changes generally concern the input variables like the telephone number, the household address, etc. and, in general, all variables whose values are known in advance, but have changed in the period among the survey waves. The telephone number is the variable that more frequently than the others changes. This is because, Istat is used to send an announcement letter before the beginning of the survey, through which it informs the sample unit about a future phone call regarding the survey and asks for its co-operation. The letter also contains a toll free number the sample units can call to ask for information or to communicate possible changes; the most frequent change is about the units’ phone number. In general, the information about this change is got before the survey starts, but quite often this is not possible and the communication arrives when the survey is in progress. To make this change a long procedure is required (the same is valid for any variables) :

- a) download the database;
- b) make this changes with a data management program (SAS or Manipula);
- c) reload the database.

Obviously, to do all these operations it is necessary to stop the fieldwork, especially for those surveys whose crew size is from morning to evening. The consequences of this are quite easily understandable.

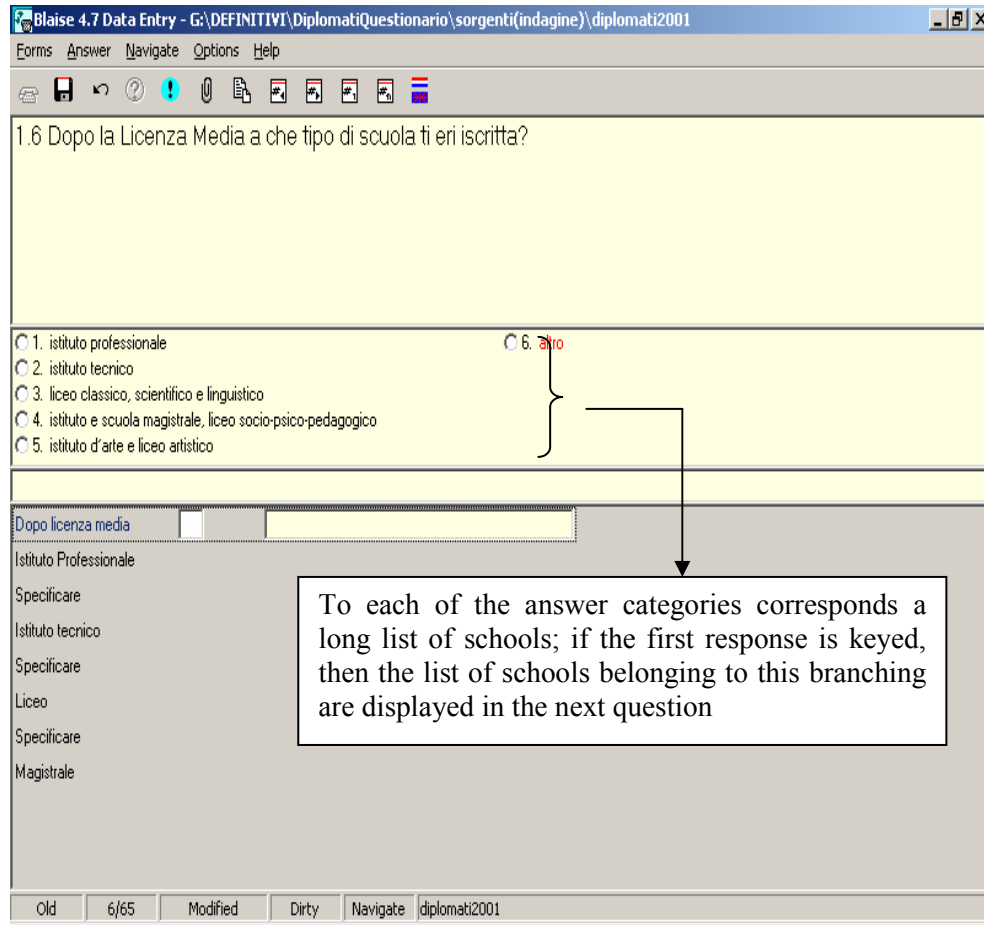
It could be hypothesised that a “browse-editing” function for the Blaise database would definitely speed up and simplify these operations.

5.5. The implementation of the item rotation

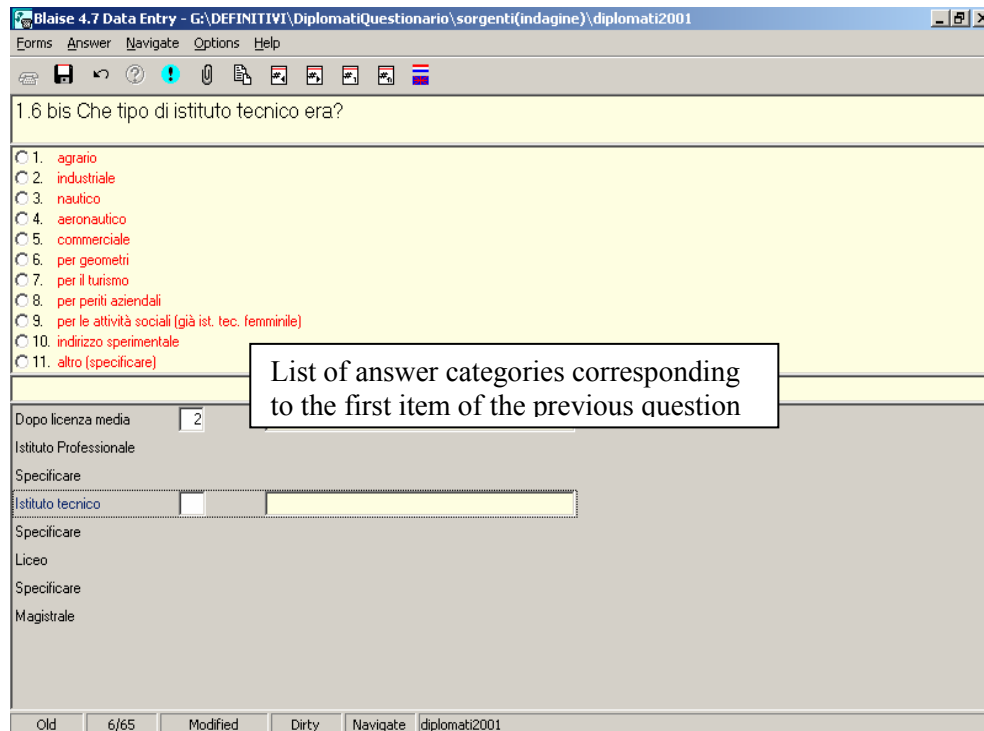
It is well known that a well designed CATI questionnaire must not contain questions with a long list of answer categories. Despite this general rule, it happens, sometimes, that a survey question does not respect this feature due to various causes.

In this case to simplify the interviewer’s job and to avoid biased results (i.e. answers concentrating on the first or last answer items) we implement different and not standardised solutions depending on the survey questions. When it is possible, we split the question in more than one in order to reach the answer step by step: this can be done only when the interview’s length is not to long (see *Example 1a and 1b*).

Example 1a How to split one question in more than one to avoid a long answering category list



Example 1b How to split one question in more than one to avoid a long answering category list



Another solution is to design the graphical layout of the question items in such a way that the answer categories are grouped according to their content (see *Example 2*).

Example 2 Answer categories grouped according to their content

Blaise 4.7 Data Entry - G:\Violenza\DB\violenza

Forms Answer Navigate Options Help Show Watch Window

REP_np25c Per quale motivo non ha denunciato questo episodio alle forze dell'ordine?
Possibili più risposte

1. **Perchè è un fatto privato**
Ho gestito la situazione da solo/ho coinvolto un amico/ è un problema da gestire in famiglia

7. Non volevo che qualcuno lo sapesse/ è un fatto privato

8. Non volevo che venisse arrestato o che avesse problemi con la polizia/carabinieri...

5. **Perchè c'è il timore...**
Ho avuto paura dell'autore della violenza/ paura delle conseguenze

6. Per vergogna, imbarazzo/pensavo fosse colpa mia

9. Non sarei stata creduta

2. **Fatto poco grave**
Non ho ritenuto che si trattasse di un reato grave/fatto insignificante/non ci ho pensato

3. **La polizia...**
Ho pensato che la polizia non avrebbe fatto niente

4. Ho pensato che la polizia non avrebbe potuto fare niente

10. **E' il mio lavoro**
E' il mio lavoro/fa parte dei rischi del lavoro

11. **Denunciato a qualcun altro**
Lo ha denunciato a qualcun altro (specificare)

12. Altro (specificare)

Firma verbale

Old 62/147 Modified Dirty Navigate violenza

All of these solutions make the implementation phase heavy as well as the use of the existing Blaise source program for the rotation.

We think that Blaise could provide a function that can manage the items rotation without the need of defining the answer categories as variables and that works in the same simple way of the instruction used for the management of multi-response questions.

5.6. Assigning a code to the dial result when the max call number is reached

Generally speaking, sample units cannot be included in the daybatch file if they have reached the maximum number of calls specified in the scheduling parameters. There is no evidence of this in their dial results, because only the last contact is registered.

As previously said, we produce reports on contact results which indicate also how many units reach the maximum number of calls. This is a very important information, because, in case of a too quick growth of this kind of units, it is necessary to understand the reasons why. It could depend from:

- a wrong setting of the scheduler parameter;
- a not well planned survey timing;
- a not updated list of sample unit;
- a wrong management of the appointments;
- etc.

Once the causes are clear, then right actions are taken to improve the situation.

To report the amount of units that have reached the maximum number of calls we have implemented a function which assigns a certain dial result considering in it

different variables: the last dial result, the number of calls and, in case of appointments, the type of appointment and the date when it was last fixed. Apart from the difficulties we met in determining this rule and the necessity to repeat this set of conditions each time a report asks for it, we are not quite sure whether or not it works properly, since the algorithm that determines the composition of the daybatch file is not well known.

Therefore, the best solution would be that Blaise assigned a code to the dial result for these cases. This should be a “not definitive” dial result, since it could change when scheduling parameters are modified.

5.7. Automatic substitution among sample units

Sometimes for a CATI survey it is not possible to use an over-sampling rate to determine the dimension of the sample size suitable to get the interviews quota. This generally happens when it is the very first survey occasion or when there is a change in the quality of the unit list that could make the over-sampling rate different from the past experiences. In all these cases the sample is divided in two sets: the base sample units set and the substitution units set. Shortly, the survey starts with the base sample units and sub-set of substitution units are added periodically and in accordance to the number of no more contactable base units, until the planned amount of interviews is reached.

To meet this requirement we implemented a software procedure that:

- a) individuates the base units to be substituted;
- b) assigns a dial result for those units who are not eligible anymore since they reached the max number of calls;
- c) determines the relative substitution units;
- d) updates the Blaise database.

This procedure is conceptually the same among the various CATI surveys, but its content varies since it depends on the substitution criteria. Probably the same result would be easily obtained if Blaise contained a procedure that loads a database containing both sets of base and substitution units and, on the basis of some parameters, automatically makes the substitution among the sample units.

6. References

Blaise for Windows Developer's Guide 4.5

Macchia S. and Murgia M. (2004). “The *in-house* strategy adopted in Istat for CATI surveys”.Q2004 European Conference on Quality and Methodology in Official Statistics. Mainz, 24-26 May.

Ryan T. P. (2000) Statistical methods for quality improvement (Second edition), Wiley, New York.

Degortes M., Macchia S. and Murgia M. (2003) “The usage of Blaise in an integrated application for the Birth Sample Survey in CATI mode” Ibuc 2003, Copenhagen, 21-23 May.

Appendix – Examples of reports for CATI surveys

Example 1. Example of report for the productivity of interviewers

REPORT 2 CUMULATED CONTACT RESULTS PER INTERVIEWER												
INTERVIEWERS												
	TOTAL		ISTATN09		ISTATN10		ISTATN11		ISTATN13			
	Abs	%	Abs	%	Abs	%	Abs	%	Abs	%		
UNTIL: 25/01/2003												
1) Completed interview												
2) Disconnected												
3) Refuse to co-operate												
4) Definitive Interruption												
5) Appointment												
6) Busy												
7) No answers												
8) Fax/Answering Machine												
TOTAL of CONTACTS												

Example 2. Example of report to monitor possible problems during the interview phase

REPORT 14 – DEFINITIVE INTERRUPTIONS PER QUESTION – Cumulated and weekly Results -				
LAST ANSWERED QUESTION	Until : 27/02/2006		Week from: 27/02/2006 to: 27/02/2006	
	DEFINITIVE INTERRUPTIONS		DEFINITIVE INTERRUPTIONS	
	Abs	%	Abs	%
Qb19.disturbi[1].disturbi				
Qb1				
Qb12				
Qb13				
.....				
TOTAL				

Example 3. Report to monitor the management of appointments when the “Route back to interviewer” option is activated

Interviewer code: 897																		
REPORT 17 ACTIVE APPOINTMENTS PER INTERVIEWER – Cumulated results																		
Date: 27/02/06																		
DATE	APPOINTEMENT TYPE						TIME SLICES											
	Exact time and exacted date	Period and day part	Week day and exacted time	Period an no time	Weekday and no time	Only day part	No date no time	TOTAL	09:00-10:59	11:00-12:59	13:00-14:59	15:00-16:59	17:00-18:59	18:00-19:59	19:00-20:59	20:00-21:00	Total	
No date																		
20060227																		
20060228																		
20060301																		
.....																		

Example 4. Report to monitor the interview's length

REPORT 8 – LENGTH OF INTERVIEW PER INTERVIEWER - Cumulated and weekly results -						
INTERVIEWERS	Until : 27/02/2006			Date : 27/02/2006		
	Completed Interviews	Average length	Difference with the average length	Completed Interviews	Average daily length	Difference with the daily average length
TOTAL						
1573						
2416						
2508						
2657						
2683						

Example 5. Report to monitor response and non response rates for each interviewer

REPORT 7 – RESPONSE AND NON RESPONSE RATES PER INTERVIEWER – Cumulated results -									
Until: 17/12/2004									
INTERVIEWERS	Response rate		Refusal rate		Interruption rate		No eligibility rate		
	Rate	Delta	Rate	Delta	Rate	Delta	Rate	Delta	
TOTAL									
1499									
1535									
1549									

