WP. 24
ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Ottawa, Canada, 16-18 May 2005)

Topic (iii): Electronic data reporting – editing nearer source and multimode collections

DATA EDITING FOR THE ITALIAN LABOUR FORCE SURVEY

Supporting Paper

Submitted by National Institute of Statistics, Italy[1]

I. INTRODUCTION

1. The European Council has approved a new Regulation (n.577/98) that aims to produce comparable information about labour market in the European Member States. The Council Regulation requires that Labour Force Surveys have to be carried out complying with both the contents and the methodologies that are set at Community level. In order to accept this Regulation the Italian Labour Force Survey (LFS) has been completely revised. The rotating panel design, the frequency of the survey and the complexity of the questionnaire are the main reasons which led to adopt a different data collection strategy, which is a combination of different *Computer Assisted Interviewing* (CAI) techniques. More exactly, CAPI (*Computer Assisted Personal Interviewing*) technique is used for the first interview, while CATI (*Computer Assisted Telephone Interviewing*) technique is used for the interviews following the first. To conduct the interviews with CAI technique, professional interviewer network has been realised.

2. As known, one of the major advantage of using CAI is the improving of data quality, especially in case of complex questionnaires. This is mainly due to the fact that the use of CAI allows to reduce large amounts of potential errors due to interviewer or to respondent. For example, routing errors are eliminated because the script automatically routes to the next questions. In addition, CAI makes it possible to check range and data consistency during the interview. As a consequence, the problem of imputing missing or inconsistent data at the post-data capturing stage is significantly reduced, even if not completely solved. A certain amount of records indeed may present a few internal inconsistencies, since some edit rules may be left unsolved during the interview. Moreover, the choice to introduce a limited number of edits, in order to do not compromise the regular flow of the interview, increases the number of records with unsolved logical inconsistencies.

3. Nevertheless, most of errors do not have an influential effect on publication figures or aggregate data, which are published by Statistical offices. On the other hand, if data have been obtained by means of a sample from the population, an error in the results due to the incorrect data is acceptable as long as this error is small in comparison to the sampling error. The question we pose is: “It is necessary to correct all data in every detail when CAI is used as data collection method?”. From a first point of view, data

[1] Prepared by Claudio Ceccarelli clceccar@istat.it and Simona Rosati sirosati@istat.it.

which are collected by means of computer-assisted methods would not need to be further corrected. On the other hand, if some edits are not solved during the interview, some records will remain internally inconsistent and may lead to problems when publication figures are calculated. Data editing helps to reduce the errors and make sure that records become internally consistent, but a caution against the overuse of query edits must be heeded.

4. It is well recognized that the ideal edit strategy is a combination of selective editing and automatic editing, although macro-editing can not be omitted in the final step of editing process. For our purpose automatic editing is especially suited for identifying all incorrect records, since performing this activity is low cost and not time-consuming. After that phase, the records are split into two stream: critical records (they may have an influential effect on aggregate data) and noncritical records (they do not affect significantly aggregate data). The critical records which are incorrect for systematic errors are imputed by a deterministic algorithm, while those which are incorrect for probabilistic errors are automatically imputed according to the Fellegi and Holt methodology (1976). The remaining noncritical records, which may be incorrect for different types of errors, could be either not edited or could be imputed automatically. If the latter solution is adopted, it is expected that software systems implementing the Fellegi and Holt method will not be able to function properly. This is also the case of the LFS. In this paper we do not give a solution to such a problem, but we are confident that useful results will be provided in the future. Suggestions, which we intend to verify, are presented in the last section.

5. The work is organised as follows: the characteristics of the LFS sample are presented in Section 2; Section 3 analyses the survey process as a whole; the edit strategy for the LFS is described in Section 4; some results about the imputation are also reported. The paper concludes with a brief discussion about the main outcomes and the issue concerning data editing in CAI surveys. For brevity all the results presented in the paper refers to 2004 fourth quarter data.

II SAMPLE DESIGN

6. The most relevant aspect about new LFS is the time continuity of reference weeks spread uniformly throughout the whole year. For this reason, all solutions adopted to realise the survey are conditioned by this characteristic. Sample design is realised taking into account Council Regulation 577/98 constraints about frequency of the survey and the representativeness of the sample.

7. The sample has two-stage of selection. The primary units are the Municipalities stratified into 1,246 group at the Province level. The stratification strategy is based on the demographic size of the Municipality. The sample scheme provides for one primary unit for each strata. The secondary units are the households and they are selected from the Municipality population Register. For each quarter 76,918 households are interviewed (i.e. sampling rate equal to 0.4%). The LFS sample is designed to guarantee annual estimates of principal indicators of labour market at Province level (NUTS III), quarterly estimates at the Regional level (NUTS II) and monthly estimates at the national level.

8. Council Regulation 577/98 indicates the quarter as the reference period about the estimates of principal indicators of labour market, so the sample of households to be interviewed is split into 13 sub groups homogeneous regarding to the size. In similar way to the old version of LFS, the sampling households follow the rotation scheme 2-2-2. For example, if we consider a sampling household that is interviewed in the third week of the first quarter of 2005, it will be interviewed again in the third week of the second quarter of 2005, then it will go out to the sample in third and fourth quarters of 2005 and, finally, it will be interviewed again in the third week of first and second quarter of 2006.

9. The weighting factors are calculated taking into account the probability of selection and the auxiliary information relating to the distribution of the population being surveyed, by sex, age (five-year are groups) and region (NUTS II level), where such external data are referred to the Register statistics.

10. The weighting strategy adopted by LFS is developed in the following there steps. In the first one, the initial weights equal to inverse of probability of selection are calculated. In the second step, non-response factors, by household size and referent person characteristic are calculated to correct initial weight from non-response effects. Finally, final weight relating to the distribution of the population being surveyed, by sex, age (five-year are groups) and region are calculated using calibration estimators methodology.

III. SURVEY PROCESS

11. Quality in statistics refers to an undefined concept. Each survey process is based on a multiple elementary operations which have an influence over the final outcome, consequently quality has progressively become a central issue in official survey designing. Starting from these considerations, the new LFS is planned to analyse the process, in terms of elementary activities. In this way, it's possible to associate each activities to a specific type of error and to define a complex quality monitoring system for all steps of the survey.

12. In order to control all elementary actions about the LFS process, the quality monitoring system is introduced and can be divided into three main phases: "preventive checks phase", regards some of preventive actions implemented before data collection, to prevent the errors; "checks during the survey phase", regards some actions for detecting the errors that can be made during the process; "*a posteriori* checks", to evaluate non sampling error using quantitative indicators linked to each elementary operation. Double strategy of interview (the first one with CAPI and the other three with CATI) call for a high level of automation on the interviews management, transmission and execution, which have to be strictly observed. Monitoring system give the possibility to put the new LFS process under control and to adjust it in "real time" in case of error (Giuliani et al, 2004).

13. Mixed mode techniques has been due to the need of interviewing the households belonging to the sample for four waves: the first interview is generally carried out by CAPI technique, the three successive interviews are carried out by a CATI technique. In particular cases, for example when the household do not have a fixed telephone, the previous scheme can be changed.

14. To carry out CAPI interviews, ISTAT decided to realise a network of 311 interviewers, one for each specific sub-regional area, directly trained and managed by ISTAT experts. Each ISTAT Regional Offices manage weekly interviewer job. Through a web connection the interviewers receive the electronic questionnaire and related changes and the information about the weekly sampling households.

15. ISTAT provided to realise a complex automatic system that has to be able to manage all information regarding the core of new LFS. In particular, the system manages the flows of data between sampling Municipalities and ISTAT, allocates the interview according to techniques, governs the flows of information in data treatment phase and so on. The information system guarantees that all activities, achieved by reliable hardware and software tools, respects all the requirements of a secure information system (data privacy, data availability and data integrity) and simultaneously, provides for monitoring system based on a series of indicators defined to plan and control all the elementary activities. (Bergamasco et al, 2004).

16. The sample Municipalities select a list of households that are the theoretical sample to be interviewed. This sample is composed by "base household" and its three "replacing households". This information are collected and managed by a software system called SIGIF (*SIstema Gestione Indagini Famiglie* – system for the management of households surveys). SIGIF distributes the sampling households according to data collection techniques (CAPI or CATI).

17. Using SIGIF, ISTAT Regional Offices allocate to the CAPI interviewer the weekly sample of households. In order to solve eventual not planned events, ISTAT Regional Offices re-assign the job between interviewer. Each interviewer try to interview the “base household” and, if it is not possible, he can replace the household with a “replacing” one chosen automatically by the CAPI system. The interviewers have about 4 weeks to carry out the interview: the first week is dedicated to fix the appointments, the second week to the interview, the following two weeks to complete pending interviews. As soon as possible, the interviewers send, using web connection to a toll-free number, the results of the weekly job to the SIGIF divided into interview and field-work data (Bergamasco et al, 2004).

18. Using SIGIF, the list of the households is sent to the private service company that has the responsibility of carrying out the telephone interview. Weekly, the service company sends to ISTAT the results of the interview and daily a set of field-work indicators. In Table 1 are reported the sample size of the LFS and the response rate by survey technique.

Table 1. Sample size and response rate by survey technique

Households			Respondent households			Response rate		
Universe	Sample	Sampling rate (%)	CAPI	CATI	Total	CAPI	CATI	Total
21,810,676	76,872	0.4	26,985	42,567	69,552	90.2	92.1	90.5

19. The European Council Regulations n. 1575/00 and n. 1897/2000 introduce other innovations, as typology of questions, type of variables (continuous or categorical), acceptance range of variables, and explore in details the concept of “unemployed person” and gives further methodological indications on data collection.

20. In order to observe these European Council Regulations and simplify the interview, electronic questionnaire has developed. Characteristics like: automatic branching, online help, interactive codification of open items using a search engine for certain key variables (such as economic activity and profession), root and coherence rules, confirmation items for waves following the first, improve significantly data quality.

21. The questionnaire is composed by a section regarding general personal information on household members. Individual questionnaire is divided into 9 sections that should be submitted to each member in working age (over 15 years); the themes treated are: labour status during the reference week, main job, second job, previous work experiences (only for the not employed persons), search of employment, registration to public and private employment office, education and training, main labour status and residence. To close the interview there are 2 section called: other information on the household, general trend of interview. The last section regards pending codification used if the interviewer do not codify some particular variables, like economic activity and profession, during the interview.

22. Rotation scheme 2-2-2 of the sampling households allows to make a panel of information. Traditionally, ISTAT produce “three months transition matrix” and “one year transition matrix” referred to longitudinal information carried out for all sampling persons. To give the possibility to reconstruct the working history of the respondent for more than one year, the longitudinal information required about the labour status has been improved.

23. European Council Regulation n. 1897/00 gives the possibility to conduct the interviews following the first asking to the sampling person to confirm the information given in the previous interview. In this way, when the labour status and/or other characteristics of the respondent does not have substantially changed we are able to reduce the time of interview. CAI technique help to realise a “confirmation

questionnaire", in which the "confirmation questions" are the fundamental nodes of interview flow. In case of the status of respondent is not changed, all information regarding a sub-flow of the questionnaire depending on certain "confirmation question" are automatically registered.

24. The LFS process and the new interviewers network represent the core of a crucial transformation of ISTAT's surveys, and has allowed, for the first time in ISTAT history, the implementation of an innovative survey management process. The attention given to data quality moved from the final product to the methodologies and the tools used to keep the whole production process under control. With this new system, the evaluation of the quality of the LFS, in terms of correctness and transparency, are set.

IV. DATA EDITING

IV.1 Error detection

25. After data captured contradictory information and item non-response are detected principally by using an automatic system based on Fellegi and Holt model of editing. It is worth noting that this operation can also be performed every week in order to notify in time possible errors due to the electronic questionnaire.

26. In editing process some edits (rules) are specified by experts with a knowledge of the subject matter of the survey. An edit expresses the judgment of some experts that certain combinations of values, corresponding to the different questions on the questionnaire, are unacceptable. The set of such edit rules is referred to as *explicit edits*. A record which fails any of the edits does need to be corrected. Conversely, a record which passes all the stated edits does not need to be corrected. The explicit edits relating to categorical variables (i.e. variables which are not subject to a meaningful metric) are implemented in SCIA (System for editing and automatic imputation), developed by ISTAT according to the Fellegi and Holt methodology (Barcaroli and Venturi, 1997). More exactly, the version of SCIA included in CONCORD (CONtrol and Data CORrection), a system designed for Windows environment, is used (Riccini Margarucci and Floris, 2000).

27. The remaining edits related to continuous variables are translated into a SAS program including if-then-else rules (SAS Institute Inc., 1999). Two different systems are used, given that a software which handle mixed data, i.e., a mix of categorical and continuous data is currently unavailable. Developing such a system is generally considered too hard, although in 1979 Sande showed how to this. He first showed how to convert discrete data to continuous data in a way that would allow solution of the error-localization problem. He then showed how to put a combination of discrete and continuous data into a form to which Chernikova's algorithm could be applied (1964, 1965). To our knowledge only recently two researcher of Statistics Netherlands proposed an algorithm based on Fellegi and Holt methods for automatic editing of mixed data (de Waal and Quere, 2003).

28. For a single variable we also distinguish two types of errors: "random error" and "systematic error". Although it is not a clear cut on their definition, we say that random errors can be assimilated to normal measurement errors, they have equal probability to occur in different variables and, for a generic variable, they are not correlated to errors in other variables. Systematic errors may be defined as non-probabilistic errors; they are generally due to some defects in surveying structure (e.g. a question wrongly specified).

29. Table 2 shows the number of explicit edits for the LFS. With regard to the SCIA system, the large number of edit rules and the complicated skip patterns of the questionnaire poses several computational problems that the entire set of edits need to be partitioned into two subsets. The problems arise since the Fellegi and Holt system runs to check the logical consistency of the entire edit system. We suppose that such a problem could be solved changing something into the software system by using more

powerful processor. As result of the stage of error detection in Table 3 are reported the proportion of incorrect records per type of error.

Table 2. Number of explicit edits

Edit	Number of edits
SCIA system	2,358
- Set 1	961
- Set 2	1,397
SAS system	142

Table 3. Erroneous records per type of error

Error	Percentage of records
Systematic	2.5
Probabilistic	2.8

IV.2 Imputation

30. The imputation of item non-responses depends on the nature of errors which generated them, although practical considerations and other issues should be considered in implementation of imputation procedures. When missing data are imputed it has been shown that deterministic imputation methods distort the distributions and attenuate variances, whereas probabilistic methods yield approximately unbiased estimates of distributions and element variances (Kalton and Kasprzyk, 1982). Nevertheless, we can reasonably assume that deterministic imputation is more suitable for correcting systematic errors, while probabilistic methods are more specific for errors generated from random error models.

31. Deterministic imputation assigns only one value, *a priori* determined, on the base of other values considered "true" by experts. On the contrary, probabilistic imputation assigns a value according to a stochastic model (e.g. a regression model) or using a donor unit which is similar to record to be imputed. A *distance function* is generally used to define similarity.

32. The Fellegi and Holt methodology can be counted among the probabilistic methods. It is concerned primarily with the stages of data editing and of imputation. Data editing process localizes errors in the record identifying a *minimal set* of variables (the smallest set), whose values can be changed in such a fashion that the resulting record would pass all the edits. Finally imputation will provide suitable values for those variables.

33. One of the major objectives of the Fellegi and Holt methodology is to retain the structure of the data. This means that univariate and multivariate distributions of survey data reflect as nearly as possible the distributions in the population. This goal is achieved by the use of *hot-deck* imputation. Hot-deck imputation consists of imputing for a variable of the current record the value recorded in the same field of some other record, which passed all the edits and which is similar in all respects to the current record. This can be done one variable at time (*sequential imputation*) or all variables at once (*joint imputation*). The former aims to preserve the univariate distributions of the variables, whereas the latter preserves the multivariate distributions.

34. To solve the error-localization problem, Fellegi and Holt showed that both explicit and implicit edits are needed. *Implicit* edits are those that can be logically derived (or generated) from a set of explicit edits. If the implicit edit fails, then necessarily at least one of the explicit edits fail. For the LFS the

current algorithm in the SCIA system can not generate the full set of implicit edits because the amount of computation needed for generation grows at very high exponential rate in the number of edits. The same limitations were found when the Fellegi and Holt system ran to check the logical consistency of the explicit edits.

35. On the other hand, it seems to be not convenient to divide further the two subsets of explicit edits: we would obtain more than three subsets, which would be not independent each from other, with additional restrictions for the imputation method. Moreover, if some records are incorrect for systematic errors, it would be better correct them by using deterministic imputation. For these reasons we adopt a procedure which is a combination of selective editing and automatic editing. After errors detection, the incorrect records are split into two stream: critical records (they are influential on aggregate data) and noncritical records (they do not affect significantly aggregate data).

36. The critical records which are incorrect for systematic errors are imputed by choosing one value *a priori* determined, while those which are incorrect for probabilistic errors are automatically imputed according to the Fellegi and Holt methodology. Currently, only probabilistic logical inconsistencies which do not depend on rules of questionnaire are automatically imputed. In order to achieve this aim 137 explicit edits are used to detect probabilistic errors and 213 edits form a complete set of edits to solve the error-localization problem, i.e., which variables change in an erroneous record to ensure that the record passes all of the edits.

37. The remaining noncritical records, which may be incorrect for different types of errors, could be left unsolved without reducing data quality. Otherwise they could be automatically imputed too. In fact, the small number of errors and the variety of failing edits make the use of probabilistic imputation acceptable. At present, the remaining incorrect records are not automatically imputed since the amount of computation needed for generating the full set of implicit edits is prohibitive. These records are imputed by a deterministic algorithm whose implementation is not only time-consuming, but need to be modified for each quarter. In order to verify that the deterministic actions are correct, all the records are checked through the edit system a second time. If they pass, then nothing else is done. If they still fail edits, then it means that either some if-then-else rules need to be adjusted or that other rules must be considered in the algorithm.

38. As result of the impact of imputation in Table 4 are reported some findings related to the records which have been corrected by sequential or joint imputation. Table 5 shows the percentage of records per number of imputations at the end of the editing process. Note that the percentage of records that are corrected for four or more variables is equal to zero. Really this percentage is equal to 0.05, which means that there are at least four variables that have been imputed in 100 records. The maximum number of imputations is equal to 16.

Table 4. Erroneous records per type of imputation

Imputation	**Percentage of records**
Joint	41.1
Sequential	58.9

Table 5. Records per number of imputations

Number of imputations	Percentage of records
0	95.5
1	3.5
2	0.7
3	0.3
4 or more	0.0

V. CONCLUDING REMARKS

39. The advantage and also the aim of imputation is to complete a data set with "full" information for all observed individuals, which reduces bias in survey estimates and – from a point of a data user – also simplifies the analysis. Nevertheless, when CAI is used the increase of data quality due to editing process is usually negligible. In addition, for large complex data set implementing an automatic system for imputing contradictory information may be too costly in terms of timeliness and resources. Therefore it seems to be unnecessary to correct data in every detail when CAI is used, while the information on error detection can play an important role in the improvement of the electronic questionnaire and the data collection. The LFS data could be checked frequently, since they are captured continuously week by week. Thus fatal errors can be identified just in time so that the errors can be eliminated. This process allows to get immediate feedback and monitoring on how interviewer are filling in the electronic questionnaire.

40. Although the impact of editing is considerably reduced by new technologies, we found that some errors may continue to be in the data and may have an influential effect on aggregate data. A combination of selective and automatic editing may be used to correct errors. Nevertheless, we also argued how it was hard implementing a system based on Fellegi and Holt methods when numerous edits are specified.

41. For the future we suggest applying the Fellegi and Holt method step by step. Initially the method is applied only for edits which fail. They are drawn from the set of the explicit edits. Then the records, which have been imputed, are checked through the edit system a second time. If they pass, then nothing else is done: all the records are correct. If the records still fail edits, the failing edits are added to the subset of edits previously selected in order to impute the incorrect records. The algorithm is reiterated until the records pass all the edits.

REFERENCES

AMENDOLA A. CAROLEO F.E., COPPOLA G. (1997) Differenziali territoriali nel mercato del lavoro e sviluppo in Italia, Celpe, Università di Salerno Discussion Paper n. 36.

APPEL M. V., NICHOLLS II, W. L., NICHOLLS, W. (1993) New CASIC Technology at the U.S. Census Bureau, ASA, U.S. Census Bureau, vol. 2, pp. 1079-1084.

BAFFIGI A. (1999) I differenziali territoriali nella struttura dell'occupazione e della disoccupazione: un'analisi con dati a livello provinciale (1981-1995). In: Struttura della contrattazione: differenziali salariali e occupazione in ambiti regionali, a cura di M. Biagioli, F.E. Caroleo e S. Destefanis.

BAKER R.P., BRADBURN N.M., JOHNSON R.A. (1995) Computer-assisted personal interviewing: an experimental evaluation of data quality and cost, *Journal of Official Statistics*, vol.11 n.4, pp. 413-431.

BARCAROLI G. (1993) Un approccio logico formale al problema del controllo e della correzione dei dati statistici, Quaderni di Ricerca, n.9, ISTAT.

BARCAROLI G., LUZI O., MANZARI A. (1996) Metodi e software per il controllo e la correzione dei dati, ISTAT, Documento interno.

BARCAROLI G., VENTURI, M. (1997) DAISY (Design, Analysis and Imputation System): Structure, Methodology, and First Applications. In: J. Kovar and L. Granquist, (eds.) Statistical Data Editing, Volume II, U.N. Economic Commission for Europe, 40-51.

BERGAMASCO S., GAZZELLONI S., QUATTROCIOCCHI L., RANALDI R:, TOMA A., TRIOLO V., (2004) New strategies to improve quality of ISTAT new CAPI/CATI Labour Force Survey, European Conference on Quality and Methodology in Official Statistics, Mainz, Germany, 24-26 May.

CHERNIKOVA, N.V. (1964) Algorithm for Finding a General Formula for the Non-negative Solutions of a System of Linear Equations, *USSR Computational Mathematics and Mathematical Physics*, 4, 151-158.

CHERNIKOVA, N.V. (1965) Algorithm for Finding a General Formula for the Non-negative Solutions of a System of Linear Inequalities, *USSR Computational Mathematics and Mathematical Physics*, 5.

COUPER M.P. (1996) Changes in Interview Setting Under CAPI, *Journal of Official Statistics*, Statistics Sweden, vol.12, n.3 pp.301-316.

DE WAAL T., QUERE R. (2003) A Fast and Simple Algorithm for Automatic Editing of Mixed Data, *Journal of Official Statistics*, Statistics Sweden, Vol. 19, No. 4, 383-402.

DECRESSING, J. E FATÀS, A. (1995) Regional Labor Market Dynamics in Europe, *European Economic Review*, n. 3, 1627-1655

FABBRIS L. (1991) Abbinamenti tra fonti d'errore nella formazione dei dati e misure dell'effetto degli errori sulle stime, Bollettino SIS, n. 22.

FABBRIS L., BASSI F. (1997) On-line likelihood controls in Computer-Assisted Interviewing, Book 1, 51° session ISI, Istanbul.

FELLEGI I.P., HOLT D. (1976) A Systematic Approach to Automatic Edit and Imputation, *Journal of the American Statistical Association*, 71, 17-35.

FILIPPUCCI C. (1998) La rilevazione dei dati assistita da computer: acquisizioni e tendenze della metodologia statistica e informatica, Sorrento, XXXIX Riunione Scientifica della S.I.S.

FUTTERMAN M. (1988) CATI Instrument logical structures: an analysis with applications, *Journal of Official Statistics*, Statistics Sweden, vol.4, n.4 pp.333-348.

FUTTERMAN, M. (1988) CATI Instrument logical structures: an analysis with applications, *Journal of Official Statistics*, Statistics Sweden, vol. 4, no. 4, pp. 333-348.

GIULIANI G., GRASSIA, M. G., QUATTROCIOCCHI L., RANALDI R. (2004) New methods for measuring quality indicators of ISTAT's new CAPI/CATI Labour Force Survey, European Conference on Quality and Methodology in Official Statistics, Mainz, Germany, 24-26 May.

GRASSIA, M. G., PINTALDI F., QUATTROCIOCCHI L., (2004) The electronic questionnaire in ISTAT's new CAPI/CATI Labour Force Survey, European Conference on Quality and Methodology in Official Statistics, Mainz, Germany, 24-26 May.

GROVES, R. M. et al. (1988) Telephone Survey Methodology, New York, John Wiley.

KALTON G., KASPRZYK D. (1982) Imputing for Missing Survey Responses, Proceedings of the Section on Survey Research Methods, American Statistical Association, 22-31.

KELLER W.J. (1995) Changes in Statistical Techonology, *Journal of Official Statistics*, Statistics Sweden, vol.11.

MASSELLI M. (1989) Manuale di tecniche di indagine: il sistema di controllo della qualità dei dati, Note e Relazioni n. 1, ISTAT.

NICHOLLS, W. L., BAKER, R. P., MARTIN, J. (1997) The effects of new data collection technologies on survey data quality. In: Lyberg, L. et al. (eds), Survey Measurement and Process Quality, J. Wiley & Sons, New York.

RICCINI MARGARUCCI E., FLORIS P. (2000) Controllo e correzione dati - Manuale utente, Istat, Roma.

SANDE, G. (1979) Numerical Edit and Imputation, Proceedings of the 42nd Session of the International Statistical Institute, Manila, Philippines.

SARIS W.E. (1991) Computer-Assisted Interviewing, Newbury Park.

SAS INSTITUTE INC. (1999) SAS OnlineDoc®, Version 8, Cary, NC: SAS Institute Inc.
