



**Economic and Social  
Council**

Distr.  
RESTRICTED

CES/1996/R.5  
15 November 1995

Original: ENGLISH

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE  
CONFERENCE OF EUROPEAN STATISTICIANS

Forty-fourth plenary session  
(Paris, France, 11-13 June 1996)

**WORK SESSION ON STATISTICAL DATA EDITING**

Note prepared by the secretariat

1. The Work Session on statistical data editing was held in Athens, Greece, from 6-9 November 1995 and was kindly hosted by the National Statistical Service of Greece. It was attended by participants from Austria, Canada, Czech Republic, Croatia, Denmark, France, Greece, Hungary, Ireland, Italy, Netherlands, Poland, Romania, Russian Federation, Slovenia, Sweden, The former Yugoslav Republic of Macedonia, United Kingdom and United States. A representative of the World Health Organisation (WHO) also attended.
2. The meeting was opened by Mrs. Konstantina Golfinopoulou, the Director General of the National Statistical Service of Greece.
3. The provisional agenda was adopted.
4. The following substantive topics were discussed at the meeting:
  - a) National reports on developments and progress in data editing;
  - b) How to evaluate the data editing process;
  - c) How to design sets of edits;
  - d) How graphics may be applied to data editing;
  - e) How new technology impacts the data editing process;

- f) What to do when an edit fails;
- g) How the computer can speed the coding process;
- h) Glossary and bibliography.
- i) Other issues concerning the preparation of Volume 2.

Agenda items b) to i) follow the subjects of the chapters of the publication. Papers prepared by Austria, Canada (5), Croatia (2), Denmark, France (3), Hungary (2), Ireland, Israel, Italy (4), Netherlands (3), Norway, Poland (3), Slovenia (2), Sweden (6), Romania, Russian Federation, United Kingdom (2) and United States (11) served as a base for discussion.

5. Mr. John Kovar (Canada) was elected Chair, and Mr. William Winkler (United States) was elected Vice-Chair.

6. Recommendations for future work are given below but other conclusions which the participants reached at the meeting on each of the above-mentioned topics are reproduced (in English only) in the Annex of this note.

7. The secretariat made available the portion of the integrated presentation of international statistical work programmes in the ECE region related to the project on statistical data editing in the Conference's programme of work (CES/1995/R.25/Add.2), and drew the attention of the participants to the goals that the Conference has identified for work in this field.

8. The Work Session recommended that the following be included in the 1995/96 programme of work of the Conference of European Statisticians:

- a) work session on statistical data editing (1996/97) to consider:
  - (i) national reports on development and progress in statistical data editing, (ii) compilation, evaluation and analysis of statistics related to the data editing process, (iii) new approaches related to data editing, imputation and coding (e.g. neural networks, new imputation methods, graphical techniques), (iv) impact of electronic data inputs on the editing process, (v) techniques used to limit over-editing, (vi) the use of external and administrative data sources in the editing process.
- b) publication of "Statistical Data Editing Methods and Techniques -- Volume 2" under the CES Standards and Studies series.

Secretariat resources: moderate.

**ANNEX**

**Summary of other conclusions reached at the Work Session on  
Statistical Data Editing in Athens, Greece  
6-9 November 1994**

**A. NATIONAL REPORTS ON DEVELOPMENTS AND PROGRESS IN DATA EDITING**

Documentation: Working papers by Ireland, Israel, Poland (2), Romania, Russian Federation, Slovenia, Sweden (2), United States, United Kingdom and an oral presentation by Canada.

1. The Work Session was informed about the state of the art in the design and implementation of data editing methods and techniques in some National Statistical Offices. It noted that the migration from mainframe to PC technology allowed direct involvement of statisticians in the data editing process as early as in the data collection phase. The Work Session noted that the organization of statistical services in many statistical agencies called for reorganization of statistical data processing. Since data editing is a labour- and money-consuming process, many statistical offices are experimenting with new data editing methods and techniques to facilitate this process and to bring data editing as close to the data collection phase as possible.

2. The Work Session took note of the experiences of some countries with selective editing. The main idea of this method is to avoid recontacting respondents with a potentially small influence on the estimate. Better results in using this method can be achieved by using a score function, that is, sensitivity to the standard error of each domain of study. The analysis of this method showed that the differences between estimates obtained through selective editing compared to the current data editing process were negligible.

3. Canada reported some experiences gained in implementation of new data editing methods for Self Administered Computerised Questionnaires in order to ensure that the respondents do not react negatively to "over-editing". The Work Session also took note that Canada is revising the publication "Quality Guidelines". The part dealing with data editing emphasizes that the primary role of editing must be to feed the cycle of continuous improvement, not for correcting errors after the fact.

4. The Work Session noted that the transition from centrally planned to market economy produced significant methodological changes in statistical services. These changes may cause confusion among local staff and should therefore be eliminated where possible by implementation of standard methods. Good examples of such an approach could be the uniform treatment of the linkage between different statistical observation units and demographic events in the life of the enterprise (e.g. fusion, division, cessation, etc.).

5. Several participants informed the Work Session about the successful use of Blaise (release 2.5) for computer-assisted personal interviewing (CAPI) and

Computer Assisted Data Input and editing (CADI). Although some limitations in the tabulation functions of Blaise have been reported, this system has become broadly used. The following major advantages of its implementation were mentioned: no rapid organizational changes in the computing environment of statistical services are necessary; it is user-friendly and allows easy control and coordination of survey processing.

6. The Work Session also discussed the necessity of interdisciplinary teamwork in the design and implementation of statistical data editing methods and techniques. Such a team should be composed of subject-matter specialists, methodologists responsible for survey design, informaticians and people managing data collection.

**B. PREPARATION OF THE PUBLICATION "STATISTICAL DATA EDITING METHODS AND TECHNIQUES", VOLUME 2**

7. The work on preparation of the publication "Data Editing Methods and Techniques", Volume 2 was steered by a group of experts composed of countries which were responsible for the coordination and pre-editing of the Chapters' contents. Canada, France, Italy, Sweden, United States and United Kingdom participated in the group.

**Chapter 1: How to evaluate the data editing process**

Documentation: Working papers by Denmark, Poland, Sweden (2)

8. The Work Session considered the documents for this Chapter to be very well prepared and the structure of the Chapter well-balanced. Bearing in mind the importance of evaluation methods of data editing for conducting the data editing process efficiently, the opinion was expressed that any other relevant papers could be invited as potential candidates for Chapter 1. It was also recommended to ascertain whether some research institute has conducted or could be influenced to conduct relevant research work. The deadline for inclusion of any new papers in the publication is mid-February 1996.

9. The discussion showed that in some National Statistical Offices there is a lack of possibilities and/or experience for the organization of training for statistical staff on how to use different data editing methods and techniques and how to evaluate their efficiency. The Work Session addressed the ECE Secretariat with the request to explore the possibility of organizing such a training course in the future.

**Chapter 2: How to design sets of edits**

Documentation: Working papers by Italy (3) and United States (3).

10. The discussion revealed that Chapter 2, as it now stands, has attained its major goal, that is to give a general overview of the current situation in National Statistical Institutes in the development and implementation of generalized data editing systems.

11. The advanced generalized systems presented in the Chapter sufficiently illustrate major trends in this area. It was pointed out that although different generalized systems exist, the same editing system ought to be used for a series of surveys to ensure their continuity. This may sometimes contradict the trend towards using the most up-to-date techniques. Those countries who presented their generalized systems in the Chapter were requested to supply Italy with information about respective references by the end of February 1996.

12. Considering the publication as a whole, it was mentioned that there are still open questions to be clarified. One of these is the clear definition of sets of edits towards their goals. Furthermore, the publication should more precisely reflect how to consider the results of the evaluation process (covered in Chapter 1) when designing sets of edits.

13. There was general agreement that problems dealing with over-editing of data sets should be considered in this chapter. The Work Session recommended that a new paper be prepared to be included in Chapter 2. Canada and Sweden volunteered to take on this task, to be completed by the end of March 1996.

14. Furthermore, it was recommended to insert into the introduction of the Chapter the notion that the design of data sets should reflect all computer platforms on which data editing is planned to be performed.

15. Many participants expressed the need for interdisciplinary team work in the design and implementation of the generalized data editing systems. Such teams should be composed of informaticians, subject-matter specialists and methodologists. The method of co-operation should also be described in the publication. The Work Session recommended addressing this issue in the introduction to the publication. Canada volunteered to prepare the introduction by the end of March 1996. Poland offered to provide Canada with the input to this task. USA volunteered to reflect this approach in the final version of the Working Paper No. 11 by the end of February 1996.

### **Chapter 3: How graphics may be applied to data editing**

Documentation: Working papers by Sweden, United Kingdom and United States (2)

16. When considering Chapter 3, the Work Session concluded that its composition and content covers the major areas of diverse possibilities for the use of graphics in data editing. It was highlighted that the use of graphics in this area shifts the data editing process more towards subject-matter statisticians. The Chapter demonstrates how the use of graphics can minimize the time necessary for data editing. It also shows how PC-oriented query graphics can improve data editing of large surveys by replacing the work with bulky printouts of detected errors with legible graphics. In addition, it describes how graphics can be used with various kinds of data editing methods.

17. The Work Session recommended the issue on 'Graphical Editing Query System of EIA' presented by USA to be included in this chapter. The participants

attended the demonstration of this product, which was followed by an interesting discussion. The Work Session expressed its appreciation of the preparation of this demonstration. USA promised to prepare the paper by the end of March 1996.

#### **Chapter 4: How new technology impacts data editing**

Documentation: Working papers by Croatia, Hungary, Slovenia, Sweden and United States (3)

18. The Work Session reached general agreement that the presented version of the Chapter is not suitable for final publication. The main reason is that the papers reflect the implementation of new technology for data collection rather than the influence of new technology on the data editing process itself. Authors of the individual papers were therefore requested to revise their contributions, highlighting the impact of technology on the data editing process. The revised version should be submitted to the United Kingdom by the end of February 1996.

19. Furthermore, it was recommended that the introduction should also be revised, to address more clearly the issues relevant to this Chapter. United Kingdom and USA agreed to formulate the revised introductory part by the end of March 1996.

20. United Kingdom will pre-edit the whole Chapter by the end of April. Canada agreed to assist in the final formulation of this paper.

#### **Chapter 5: What to do when an edit fails**

Documentation: Working papers by Canada (3), the Netherlands (3) and Sweden.

21. The discussion reconfirmed that there is still a tremendous potential for the reduction of costs of data editing via the implementation of relevant selective editing methods, and therefore the topic of this Chapter is one of the most important issues of the publication. The content of the Chapter is also valuable also because the presented methods reflect the most recent developments in this area.

22. Apart from the general introduction, all the papers submitted for consideration under this Chapter had already been discussed at the 1994 Work Session in Cork and reflected the outcome of this discussion. The Work Session expressed its satisfaction with the level of preparation of the Chapter and adopted it.

23. Netherlands presented the state-of-art of the currently developed system for edit and imputation in economic surveys. This system is based on Chernikova's algorithm and, in final form, it will also encompass heuristics enabling a reduction of computing time. The Work Session expressed its interest in this project and requested the Netherlands to keep the Work Session informed of further progress in the development of this project.

**Chapter 6: How the computer can speed the coding process**

Documentation: Working papers by Austria, Canada, Croatia, Hungary, the Netherlands and France (3).

24. The Work Session agreed that Chapter 6 will consist of three major parts, i.e. the theoretical explanation of automatic coding, major methods used for computer-assisted coding and different applications of automatic coding. France was requested to prepare the revised version of the introduction by the end of January 1996, clearly reflecting which papers belong to the above-mentioned areas. The revised version should be sent to the members of the Steering Group.

25. The Work Session agreed to include papers on automatic coding presented in the Work Session on Statistical Data Editing in 1993 in Sweden. It requested France to contact the authors of respective papers, asking them to revise their papers in line with the main goal of Chapter 6, and requesting them to provide France with the papers in an electronic format.

26. Furthermore, it was agreed that France will complete a paper describing the system SICORE and will include it into the Chapter.

27. It was agreed that France will pre-edit the whole Chapter by the end of February 1996 and send it to the Steering Group.

**Chapter 7: Glossary and bibliography**

Documentation: Working paper by United States

28. The Work Session expressed the opinion that the Glossary is out-dated and should be substantially revised. USA agreed to take over the responsibility for this task. United Kingdom and Canada volunteered to cooperate with USA in this work, to be completed by the end of April 1996.

29. An important source for this update are the papers contributed to Volume 2. The Work Session therefore requested the coordinators of individual Chapters to ensure that the authors will identify the terms and definitions that should be included in the Glossary by the end of January 1996. The Chapter coordinators should send these contributions to USA by the end of February 1996.

30. The Work Session recommended to shorten the Glossary to the extent suitable for publication and to include the Glossary in Volume 2. It was also recommended that the Glossary be prepared as a list in sequential order without any content-oriented subdivision.

31. The Work Session considered the form of References to the publication. It was recommended that the References reflect only Volume 2, i.e. the References on statistical data editing published as part of Volume 1 will not be updated by References needed for Volume 2.

**Other issues concerning the preparation of Volume 2.**

32. The Work Session considered it necessary to prepare an introduction to Volume 2. Canada, Sweden and United States volunteered to prepare it by the end of March 1996.

33. In the light of the introduction, the order of the individual Chapters will finally be specified. The Work Session also recommended to change the titles of individual Chapters to better reflect that Volume 2 contains guidelines and recommendations rather than direct instructions on how to solve individual tasks.

34. The Work Session recommended that the Conference authorize that Volume 2 be submitted for publication in the middle of August 1996, and that the deadlines for all activities needed for its finalization should be scheduled in this respect. The ECE Secretariat is responsible for the final layout and publication of Volume 2.

35. Bearing in mind the deadlines, the Work Session strongly recommended the reasonable limitation of all new inputs for the publication. It also recommended to shorten individual contributions to maximum extent possible.

36. The Work Session agreed to proceed with the editing of Volume 2 in two steps: the Steering Group will be responsible for the content-oriented editing of individual Chapters and it should be finished by the end of April. The ECE Secretariat will be responsible for the English editing, the final deadline for which is the end of June 1996.

37. The Work Session decided that the individual contributions to Volume 2 should be submitted on diskette in Word Perfect 5.2 format (textual and graphical part). In those cases where it is not possible to convert the graphical part into the Word Perfect format, it should be arranged as an annex to the textual part and submitted to the ECE Secretariat as hard copy. The leaders of the individual Chapters will be responsible for the submission of both diskettes and/or hard copies to the ECE Secretariat by the middle of May 1996.

**C. FUTURE WORK**

Documentation: Working papers by Italy, Canada and Norway

38. The participants discussed topics for future work. The following areas were mentioned:

- a) compilation, evaluation and analysis of statistics related to the data editing process;
- b) new approaches related to data editing, imputation and coding (e.g. neural networks, new imputation methods, graphical techniques);
- c) impact of electronic data inputs on the editing process;
- d) techniques used to limit over-editing;

- e) the use of external and administrative data sources in the editing process.

39. Contributions to these topics by the end of May 1995 were offered by the following countries: (a) Sweden, UK and USA; (b) Canada, Italy, Norway and UK; (c) Greece, Ireland, Netherlands and Slovenia; (d) Netherlands and UK; (e) Denmark, Romania and USA.

**D. OTHER BUSINESS**

40. The Work Session was informed that Statistics Netherlands preliminarily offered to host the next meeting of the Work Session in Voorburg from 4-7 November 1996.

41. The Work Session expressed appreciation to the National Statistical Service of Greece for having hosted this meeting and for the excellent organization and support in all Work Session activities.