

CONFERENCE OF EUROPEAN STATISTICIANS

UN/ECE Work Session on Statistical Data Editing
(Cardiff, United Kingdom, 18-20 October 2000)

INFORMATION NOTE TO PARTICIPANTS

The UN/ECE Work Session on Statistical Data Editing will be held at the Cardiff International Arena, Mary Anne Street, Cardiff, United Kingdom, starting on 18 October at 9.00 a.m.

I. PROVISIONAL AGENDA OF THE MEETING

1. Adoption of the agenda.
2. Designation of officers.
3. Management and evaluation of editing and imputation procedures
Invited papers by Denmark/Norway, Germany, United Kingdom and United Kingdom/United States; supporting papers from Canada, Denmark, Estonia, Ireland, Norway, Romania, and United States.
4. Propagation of knowledge to users
Invited papers by Canada, France, Italy and Sweden; supporting paper from Hungary.
5. New techniques and tools for editing imputations
Invited papers by Canada, Netherlands, Spain and United Kingdom; supporting papers from Canada, Ireland, Kazakhstan, Netherlands, Russian Federation, Switzerland and United States.
6. Evaluation of efficiency of statistical data editing
Invited paper by Prof. Svein Nordbotten (Norway); supporting papers from Canada, France, Netherlands, Sain, Sweden and United Kingdom.
7. Future work.
8. Adoption of the conclusions reached at the Work Session.

II. GENERAL INFORMATION

9. The Conference of European Statisticians, at its 1999 plenary session, included this Work Session in its meeting programme (see CES/1999/30/Add.2, programme element 2.2). The Work Session will be held in English only.
10. All documentation for the meeting will be made available at the following Internet address:

<http://www.unece.org/stats/documents/2000.10.sde.htm>

Hard copies of documents will be sent out to participants only if requested (please indicate your request on the Registration Form, or send an e-mail to *diane.serikoff@unece.org*). The UN/ECE Secretariat will inform participants by e-mail when documents become available on the Internet.

11. The agenda items will be considered based on two main categories of papers, **invited papers** and **supporting papers**. The organisers of the meeting have contacted a small number of countries and asked them to prepare invited papers for specific items. The invited papers will provide the main basis for discussion. However, issues raised in the other papers that countries contribute to the meeting (supporting papers) will also be discussed under that item. The deadline for the receipt of all papers is 31 May 2000.

12. Consideration of individual agenda items will be led by Session Discussants. The sessions will start with the presentation of the invited papers, followed by a brief question period. The Discussants will then introduce the supporting papers. The authors will have an opportunity to comment briefly on their papers and to answer any questions. The Session Discussants and authors of both invited and supporting papers are encouraged to keep in close contact in order to better prepare the discussion and to take into account the requirements of the Session Organizers. The Session Discussants and their relevant contact information are provided below.

III. EXPLANATORY NOTES TO THE AGENDA

13. The meeting will focus on the substantive topics listed in the Programme of Work of the Conference of European Statisticians under the Programme activity 2: *Technical infrastructure and other cross-cutting issues*, Programme element 2.2: *Statistical data collection and processing* (see CES/1999/30/Add.2).

Agenda item 3: Management and evaluation of editing and imputation procedures

Discussant: Leopold Granquist (Statistics Sweden)

Addr.: Statistics Sweden, P.O. Box 24300

S-10451 Stockholm, Sweden

Tel: +468 5069 4443

Fax: +468 5069 4599

e-mail: leopold.granquist@scb.se

Invited papers:

- *How to measure the effect of data editing* (Birger Madsen, Denmark, and Leiv Solheim, Norway);
- *Multivariate analysis of editing and imputation quality indicators* (Paul Smith, United Kingdom, and Paula Weir, United States);
- *Issues in evaluation of imputation performance* (Ray Chambers, United Kingdom);
- *The planning of data editing* (Elmar Wein, Germany).

14. The session will review various aspects on measuring quality of editing and imputation procedures. The invited papers will consider different ways to measure the effect of data editing and information management for planning editing processes.

15. The paper by Denmark and Norway discusses how the effect of data editing can be measured with respect to: (i) groups defined by the row and column variables in tables, and (ii) groups defined in other ways, e.g. by dividing the population into sub-populations of particular interest. A formula for measuring the effect of data editing is presented. Another issue is the actual process of editing the units that have the largest impact on the results.

16. The aim of the joint paper by United Kingdom and United States (by Paul Smith and Paula Weir) is to look at the main dimensions of variation in “quality” as determined by the edit and imputation indicators. Consequently a set of “most important indicators” will be defined and the analysis is focused on the best and worst domains in a survey as demonstrated by these dimensions of variation. The approach is to obtain data on quality indicators using one survey in the UK and one in the US; to put all the indicators into a dataset (one row containing all the indicators for a domain), and to find the principal components of this dataset; and finally add other quality indicators not related to editing and imputation to this dataset, in particular sampling error, and to repeat the principal component derivation.

17. The paper by United Kingdom (by Ray Chambers) considers imputation for “missingness” in survey databases, where missingness is caused by either (a) nonresponse by the surveyed units, or (b) “incorrect” response by the surveyed units. The goal is the specification of evaluation criteria for edit and imputation schemes that will be used in EUREDIT.

18. In the German paper, the focus is on information management for planning editing processes and it analyses data editing specifications, output used for performance of data editing, and verification and optimisation during the performance of data editing.

19. Based on these and other contributions, the meeting is expected to consider the methods to evaluate the editing and imputation procedures as a prerequisite for the efficient management of data editing. The aim of the discussion would be to identify good practices on how to use information from the editing stage in optimising resources devoted to editing and for improving other stages of the survey process. Another aim would be to develop proposals on how to standardise the evaluation methods to form an objective basis for comparing different data editing applications.

20. The principles for constructing data for monitoring and evaluation of the editing process can be analysed in more detail, e.g., how to construct and use benchmark data and proposals for concrete benchmarks of data editing indicators. The discussion can also deal with using data quality measures to manage the editing and imputation process, such as how to define and measure the quality of the process, the relation between process quality and data quality, and how to quantify the level of quality improvement.

Agenda item 4: Propagation of knowledge to users

Discussants: Giulio Barcaroli (ISTAT, Italy) and Claude Poirier (Statistics Canada)

Addr.: Mr. Giulio Barcaroli, Senior Researcher, ISTAT, Methodological Studies Service
Via C. Balbo 16, 00184 ROME, Italy
Tel.: +3906 4673 3218
Fax: +3906 4673 3221
e-mail: barca@istat.it

Mr. Claude Poirier, Chief, Generalized System Methods Section, Statistics Canada
120 Parkdale Ave., OTTAWA K1A 0T6, Canada
Tel:+1-613 951 1491
Fax:+1-613 951 5711
e-mail: poircla@statcan.ca

Invited Papers:

- *Metadata about editing for end users of statistics (Prof. Svein Nordbotten, Sweden);*
- *Propagation of knowledge on statistical quality in the case of business surveys (Pascal Rivière, INSEE, France);*

- *A prototype knowledge base on data editing and imputation (Claude Poirier, Statistics Canada);*
- *Data editing methods and techniques: knowledge to and from users (Orietta Luzi, Antonia Manzari, ISTAT, Italy).*

21. This session includes four invited papers which consider the transfer of knowledge among users of edit and imputation techniques. While the first contribution from Sweden concentrates mainly on *external* users' needs for information and how they should be met, by proposing a framework for informing them of the quality of survey data, the others are mostly directed towards the knowledge of users *within* Statistical Institutes. In particular, France's contribution addresses statisticians who manage business surveys, staff in charge of manual data editing, people responsible for dissemination and researchers; it details the available tools and documentation for business surveys and explains how they can be disseminated and reused. Canada's contribution proposes a public *knowledge base* containing a repository of data editing and imputation techniques; it also includes the presentation of a prototype base, starting from an existing glossary and an evaluation of current software. Finally, Italy's contribution deals with an analysis of information flows between the Statistical Institute internal users (mainly statisticians who manage surveys) and methodologists who are experts in data editing techniques, in the different steps of a *data editing method life cycle*.

22. Based on these papers and other contributions, this session is expected to consider the transfer of knowledge on the editing and imputation process to internal and external users. On the one hand, the propagation of knowledge concerns technical issues, i.e. the methods to capture, store and make available to users information about the data editing process and its impact on data. On the other hand, this topic relates to educating users in editing and imputation methods. Educating users requires the development of training materials, teaching subject-matter statisticians modern editing and imputation procedures, convincing staff of the advantages of automatic editing procedures, development and maintenance of a knowledge base on data editing methods which would serve as a repository of data editing terminology, methods and practices, etc. The exchange of experience between the staff working on different projects within the same office, as well as between statistical offices is also very important in this respect.

Agenda item 5: New techniques and tools for editing/imputation

Discussant: Bill Winkler (U.S. Bureau of the Census)

Addr.: Statistical Research Division, Bureau of the Census, Washington, D.C. 20233-9100,
U.S.A.

Tel: +1-301 457 4729

Fax: +1-301 457 4726

e-mail: bwinkler@census.gov

Invited papers:

- *Canadian Census Minimum Change Donor Imputation Methodology (Mike Bankier, Canada);*
- *New developments in automatic edit and imputation at Statistics Netherlands (Ton de Waal, the Netherlands);*
- *Analysis and Quality Control from ARIMA modelling (Pedro Revilla, Spain);*
- *Using new tools to improve the efficiency of the editing process in business surveys (Pam Davies, United Kingdom).*

23. The invited papers under this topic will provide information about new developments in data editing and imputation.

24. The Canadian paper presents the developments in the nearest-neighbour imputation methodology (NIM) for the 2001 Canadian Census (CANCEIS). While the focus of processing in 2001 will still be

more on qualitative rather than quantitative variables, CANCEIS also has the capacity to perform imputation for numeric variables. In the presentation, the focus will be on the recent extensions of NIM to numeric editing and imputation. The initial version of CANCEIS should be available before October and will have been given to ISTAT for inclusion in the EUREDIT project testing programme. Current plans indicate that implementations of NIM will be used in the upcoming Brazilian and Italian Censuses plus the Italian Labour Force Survey.

25. The paper from the Netherlands describes new algorithms and prototype computer programmes for automatic edit and imputation developed by Statistics Netherlands in collaboration with other institutes. The recently developed algorithm for automatic editing can be applied to a mix of continuous and discrete data and it is based on generating small sets of implied edits a large number of times. The new prototype software is based on weighted automatic interaction detection that is used to fit a tree-based model to data. The developed software can be applied to impute for missing values of both discrete and continuous data.

26. The paper from Spain describes the data characteristics for analysis and quality control from ARIMA modelling. ARIMA modelling is used to estimate a set of characteristics of a short-term indicator (e.g. the index of industrial production), as trends, seasonal variations, cyclical oscillations, unpredictability, deterministic effects (as a strike), etc. Thus, for each sector and product a vector of values corresponding to the above mentioned characteristics that can be used for data editing is constructed.

27. The paper from the United Kingdom presents new tools to improve the efficiency of the editing process in business surveys. The Office for National Statistics are undertaking a methodological review of data collection and validation activities for business surveys. The review is expected to lead to efficiency savings with no adverse impact on quality. Research is focusing on four main areas: data validation, data editing, questionnaire design and quality measurement. All four main topics include some research that is based on the development of new tools and techniques, or that is the first application of such tools in the business survey context.

28. The discussion on these and other papers contributed to the meeting is expected to touch upon the operational feasibility and the organizational issues associated with the implementation of new methods and techniques. The meeting could also try to identify potential areas for further research on data editing methods and tools.

Agenda item 6: Evaluation of efficiency of statistical data editing

Discussant: John Kovar (Statistics Canada)

Addr.: 120 Parkdale Ave., Ottawa, Canada, K1A 0T6
 Tel: +1-613 951-8615
 Fax: +1-613 951-5711
 e-mail: kovar@statcan.ca

29. The discussion on the evaluation of efficiency of statistical data editing is a continuation of consideration of this topic at the previous Work Session. It will be based on the methodological material "Evaluating efficiency of statistical data editing" prepared by prof. Nordbotten and will consider the practical applications of this theory based on contributions by Canada, France, the Netherlands, Spain, Sweden and United Kingdom. The aim of the discussion is to propose recommendations/guidelines on how to implement this theory in statistical practice.

Agenda item 7: Future work

30. Under this agenda item, the Work Session is expected to discuss issues of potential interest to countries for future work on statistical data editing.

IV. DEMONSTRATIONS

31. Since only a very limited amount of time can be set aside during the meeting for demonstrations, they will have to be made on the margins of the meeting (e.g. during the lunch break and after the afternoon session ends). To facilitate planning, participants wishing to give a demonstration should inform the UN/ECE secretariat well in advance of the meeting and also indicate the requested facilities and equipment for the demonstrations. It should be noted that there will be time for only a limited number of presentations at the meeting.

V. ORGANIZATIONAL INFORMATION

General information

32. The Work Session will be hosted by the Office for National Statistics, United Kingdom (ONS) at the Cardiff International Arena (CIA), Mary Ann Street, Cardiff, CF1 2EQ, from 18-20 October 2000. Lunch will be provided at the conference venue.

33. To obtain general information about Cardiff (how to get there, what to see, where to eat etc.) we suggest a visit to the following web-site: www.cardiffmarketing.co.uk.

Travel

34. From Cardiff Airport there is a bus that runs to the centre of Cardiff. This takes approximately 45 minutes, and buses leave every hour. You will need the X91 service up to 18:00 and the 345 service after 18:00. There is also a taxi stand just outside the airport (cost = £16 per car and £24 per 7 seater mini-bus).

35. From Heathrow Airport (London) there is the Heathrow Express rail link which runs directly to Paddington Station. This link runs every fifteen minutes and takes approximately fifteen minutes.

36. From Paddington Station there are trains running every hour to Cardiff Central, journey time - two hours. Cardiff Central Train Station is just a five minute walk away from the CIA.

Accommodation

37. **Participants are asked to make their own hotel reservations.** ONS has been able to obtain special prices at some of the hotels listed below (marked *). For the discounted hotels you will need to make a reservation through our booking agent:

Expotel - Tel: +44-161 947 4595

Ask for the Government Desk

QUOTE: Expotel Code - OFFNATNEW

Cost Centre Code - SD10

You will need to settle your own account on departure.

38. Below are also details of hotels which can be booked directly. If there are problems with hotel reservations, please contact Kath Evans at ONS on:

Tel: +44 1633 812835

Fax: +44 1633 652668

e-mail: kath.evans@ons.gov.uk

Higher Rates	
*Hilton Cardiff, Kingsway, Cardiff (5 star) Tel: +44 29 20646300 15 minutes walk to CIA	£95.00
*The Thistle Cardiff, 1 Park Place, Cardiff (4 star) Tel: +44 29 20383471 10 minutes walk to CIA	£86.00
Jurys, Mary Ann Street, Cardiff (4 star) Tel: +44 29 20341441 2 minutes walk to CIA	£85.00
*Forte Post House, Castle Street, Cardiff (3 star) Tel: +44 29 20388681 15 minutes walk to CIA	£60.00
Medium Rates	
*The Big Sleep Hotel, Harlech Court, Bute Terrace, Cardiff (2 star) Tel: +44 29 20636363 2 minutes walk to CIA	£53.50
*The Grand Hotel, Westgate Street, Cardiff (2 star) Tel: +44 29 20373317 10 minutes walk to CIA	£49.50
*Campanile Hotel, Caxton Place, Pentwyn, Cardiff (2 star) Tel: +44 29 20549044 15-20 minute bus journey (buses every 20 minutes)	£39.50 Room only
Lower Rates	
Courtlands Hotel, 100 Newport Road, Roath, Cardiff (2 star) Tel: +44 29 20497583 20 minutes walk to CIA	£28
St Hilary Hotel, 144 Cathedral Road, Pontcanna, Cardiff (2 star) Tel: +44 29 20340303 5 minute bus journey (buses every 10 minutes)	£18 std/ £25 en suite
The Briars, 128 Cathedral Road, Pontcanna, Cardiff (2 star) Tel: +44 20 20340881 5 minute bus journey (buses every 10 minutes)	£18 std/ £25 en suite

Currency

39. Some exchange rates between the Pound and other currencies are given below (valid as of 1 May 2000):

1 Pound = 1.7039 Euro

1 Pound = 10.1307 FIM

1 Pound = 11.1766 FRF

1 Pound = 2.2992 CAD

1 Pound = 1.5530 USD

1 Pound = 3299.15 ITL

1 Pound = 2.6647 CHF

Conference Dinner

40. There will be a conference dinner at Cardiff Castle on Thursday 19 October at 7.00 p.m. Please indicate on the booking form whether you would like to attend.

Excursion

41. It is planned to organise a whole day or half day excursion to the Museum of Welsh Life (part of the National Museum of Wales) on Tuesday 17 October (at a cost of appr. £10 per person), if there is sufficient interest. Please indicate on the booking form if you would like to participate.

Additional Information

42. Any questions concerning organisational arrangements for the Work Session should be addressed to:

Paul Smith
Tel: +44 1633 813436
Fax: +44 1633 813166
e-mail: paul.smith@ons.gov.uk

