

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing  
(Prague, Czech Republic, 14-17 October 1997)

Item 3 of the provisional agenda

**NATIONAL REPORT ON DEVELOPMENT AND PROGRESS IN STATISTICAL  
DATA EDITING**

Submitted by Statistics Norway <sup>1</sup>

---

<sup>1</sup> Prepared by Frank Foyn.

## **A. General project: Statistical handbook in data editing**

1. Statistics Norway has started a project on the evaluation of our data processing system and to propose measures for standardization and rationalization of methods. The intention is to develop a general handbook on the treatment of data from the different kinds of surveys. The project covers the activities from the moment data arrive at our office until they are ready for publishing. This encompasses our routines for receiving data, file-registration, manual and EDP-based checks and controls, corrections, editing and imputing. Our goal is to improve these processes to make them more efficient. We are specifically interested in measuring the effects of all the corrections and edits and in seeing whether it is possible to substitute manual revision work with more statistical editing methods. As part of the project, we will identify 2 to 3 specific statistical surveys and study in greater depth how we can use more automatic methods and neural network in the statistical process.

2. This project will encompass both personal-based and enterprise-based surveys, short-term statistics and structural statistics, questionnaire-based surveys and administrative information. We will also take into account statistics where the quality of individual microdata is important (research studies etc). and statistics where high quality only at the aggregate level is necessary.

3. The project will not include data collection methods, survey methods, questionnaire design etc., even though these aspects are important for both the quality of data, and the publishing process (publications, databases, storing etc.).

## **B. Population census**

4. The reference date for the combined Population and Dwelling census is planned to be November 2000. The population part will be fully register-based and several registers will be used in addition to the Population Register. In the preparation process there will be quality controls between the registers for a sample of individuals for a number of variables.

5. The dwelling part will be a paper-based total survey on all dwellings. There are plans to establish a Norwegian Dwelling register, but this will not be fulfilled within the time limit of the census. The procedures for editing and imputation of data are for the time being not at all discussed in detail. The census for 1990 was based on data from administrative registers and survey data from a sample of the population.

## **C. Use of register data in the Labour Force Survey (LFS)**

6. For many years Statistics Norway has used registers as the target population for sampling surveys. Drawing a sample from a register-based target population means using a smaller sample to achieve an acceptable quality on the estimates. We also use register information in the data collection and estimation processes.

#### **D. Use of register data for replacing questions**

7. Values on variables in the register may replace questions in a survey. The condition is that the definition of the variable, the quality and the timeliness of the information be acceptable. The advantage of using register information is that the response burden is reduced and that the cost of the data collection is lower since both the registration and coding of data is superfluous. The quality may also be higher if the variables in the register are more consistent than in the questionnaire-based surveys and this is often true. On the other hand, errors in register data will often be systematic and then cause bias in the dataset.

8. In the Norwegian LFS, register data is extensively used, but in some cases they only partly replace survey questionnaire. This is the case for the variable «Highest education». Before undertaking the interview survey, the sample in LFS is linked to the Education register and we identify those persons without a registered education in the register. Those persons will then be asked to participate in the survey. As timeliness is a problem in the Education register, we ask the others if they have completed any education since October last year. If yes, they should answer what kind of education this is.

#### **E. Use of register data in data editing**

9. Register data cannot for several reasons replace questions of employment in the Norwegian LFS. But even in these cases we can use information from the register in data editing. This is done for the classification of activity branch (NACE). In LFS the coding of NACE is done on the basis of the establishment's name, address and activity. The sample in LFS is linked to the Employment register where we have information on the establishment in which the person is employed. If the establishment from the two sources seems to be identical, the NACE-class in the register is automatically chosen. The NACE-class in this register comes from our central Enterprise and Establishment register which is based on detailed information about the establishment and this information is considered of better quality than in the LFS in general. In the LFS the answers often are vague regarding the establishment in which the person is employed.

In cases where the information on the Employment register is not available, we can find the establishment's name, and thus NACE-class, by direct search in our central Enterprise and Establishment register.